

# IOWA STATE UNIVERSITY

## Digital Repository

---

Retrospective Theses and Dissertations

Iowa State University Capstones, Theses and  
Dissertations

---

1997

# Econometric analysis of measurement error in panel data

Elizabeth Martha Paterno

*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Economics Commons](#), and the [Statistics and Probability Commons](#)

---

## Recommended Citation

Paterno, Elizabeth Martha, "Econometric analysis of measurement error in panel data " (1997). *Retrospective Theses and Dissertations*. 12020.

<https://lib.dr.iastate.edu/rtd/12020>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# UMI

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA  
313/761-4700 800/521-0600



**Econometric analysis of  
measurement error in panel data**

by

**Elizabeth Martha Paterno**

A dissertation submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
**DOCTOR OF PHILOSOPHY**

Co-majors: Economics; Statistics

Major Professors: Yasuo Amemiya and J. Arne Hallam

**Iowa State University**

**Ames, Iowa**

**1997**

**UMI Number: 9814681**

---

**UMI Microform 9814681**  
**Copyright 1998, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized  
copying under Title 17, United States Code.**

---

**UMI**  
**300 North Zeeb Road**  
**Ann Arbor, MI 48103**

Graduate College  
Iowa State University

This is to certify that the Doctoral dissertation of  
Elizabeth Martha Paterno  
has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

**Co-major Professor**

Signature was redacted for privacy.

**Co-major Professor**

Signature was redacted for privacy.

**For the Co-major Program**

Signature was redacted for privacy.

**For the Co-major Program**

Signature was redacted for privacy.

**For the Graduate College**

## TABLE OF CONTENTS

<b>GENERAL INTRODUCTION . . . . .</b>	<b>1</b>
Measurement Error . . . . .	2
Errors-in-Variables Models for Panel Data . . . . .	4
Moment Structure Analysis . . . . .	7
Covariance Structure Analysis . . . . .	8
Mean and Covariance Structure Analysis . . . . .	9
Dissertation Organization . . . . .	10
References . . . . .	11
<b>ANALYSIS OF PANEL DATA USING A RANDOM EFFECT     ERRORS-IN-VARIABLES MODEL . . . . .</b>	<b>15</b>
Abstract . . . . .	15
Introduction . . . . .	15
General Approach and Identification . . . . .	19
Estimation . . . . .	24
Asymptotic Properties . . . . .	25
Simulation Study . . . . .	28
IV/GMM Estimation versus ML Estimation . . . . .	29
Finite Sample Properties of the ML Estimator . . . . .	33
Summary . . . . .	36
References . . . . .	36
<b>RANDOM COEFFICIENT REGRESSION WITH ERRORS IN VARIABLES . . .</b>	<b>38</b>
Abstract . . . . .	38
Introduction . . . . .	38
Model and Identification . . . . .	40
Estimation . . . . .	44

Simulation Study . . . . .	46
Summary . . . . .	53
References . . . . .	54
<b>GENERAL CONCLUSIONS . . . . .</b>	<b>56</b>
<b>APPENDIX 1: ADDITIONAL SIMULATION RESULTS FOR THE RANDOM EFFECT ERRORS-IN-VARIABLES MODEL . . . . .</b>	<b>57</b>
<b>APPENDIX 2: ADDITIONAL TABLES FOR RANDOM COEFFICIENT REGRESSION WITH ERRORS IN VARIABLES . . . . .</b>	<b>65</b>
<b>ACKNOWLEDGMENTS . . . . .</b>	<b>70</b>



## GENERAL INTRODUCTION

A panel data set consists of measurements taken from several individuals over time. These repeated measures can often be made where single measurements can be made, unless the measurement process is destructive (Crowder and Hand, 1990). The availability of panel data has facilitated empirical research on a wide variety of areas in economics. While one-time retrospective interviews tend to be subject to recall error, panel surveys tend to generate more reliable data since the time lapse between the interview and the time periods to which the questions refer is reduced (Solon, 1989). Moreover, panel data sets allow researchers to analyze a number of economic questions that would not otherwise be possible to address using only cross-section or only time-series data sets. For example, one analytical benefit of panel data relative to cross-sectional data is that panel data make possible the measurement and analysis of micro-level dynamics. In addition, the magnitude of omitted variables may be controlled and accounted for using panel data sets.

Unlike a data set that is obtained by pooling several cross-sectional data sets (each involving a different set of individuals) taken at different time points, a panel data set has the special characteristic that some observations may not be stochastically independent. In particular, measurements taken from the same individual are likely to be correlated in some way. It is this characteristic that demands a special set of analysis techniques. To account for the individual variability, statistical models, such as random effect and random coefficient models, that contain individual random components are often used in panel data regression analysis. The random effect model assumes that

$$\begin{aligned} Y_{it} &= \alpha_i + \beta' x_{it} + e_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \\ \alpha_i &\sim (\mu_\alpha, \sigma_{\alpha\alpha}), \\ e_{it} &\sim (0, \sigma_{ee}), \end{aligned} \tag{1}$$

where  $\alpha_i$  and  $e_{it}$  are independent of each other, and  $\alpha_i$ ,  $i = 1, \dots, n$ , represent random unobserved individual effects possibly correlated with the  $k \times 1$  vector of explanatory variables,  $x_{it}$ , and  $\beta$  is a  $k \times 1$  vector of unknown parameters. The average intercept for this population of individuals is  $\mu_\alpha$ , and  $\sigma_{\alpha\alpha}$  is the variability in the response  $Y_{it}$  due to the individual differences. On the other hand, the random

coefficient model is

$$Y_{it} = \beta_i' \begin{pmatrix} 1 \\ \mathbf{x}_{it} \end{pmatrix} + e_{it},$$

$$\beta_i \sim (\beta, \Phi), \quad (2)$$

where  $\beta_i$  is independent of  $e_{it}$ ,  $\beta$  is the average coefficient, and  $\Phi$  represents the individual variability in all coefficients. Clearly, (1) is a special case of (2), and the random coefficient model (2) provides a very flexible way of accounting for individual variability in the panel data regression.

Most econometric textbooks (e.g., Johnston, 1984; Judge *et al.*, 1985; Greene, 1993) discuss the existing methods for analyzing models (1) and (2). Chamberlain (1984), Hsiao (1986), Matyas and Sevestre (1992), and Baltagi (1995b) also contain extensive discussions on the econometric analysis of panel data. Survey articles on the same topic include Maddala (1987) and Baltagi and Raj (1992). Likewise, special issues of journals have also been devoted to this subject (Heckman and Singer, 1982; Baltagi, 1995a).

## Measurement Error

The existence of measurement errors in most economic data has long been recognized (see, e.g., Morgenstern, 1963). Not only do measurement errors arise from the survey process in the form of transcription, transmission or recording errors, variables may also be said to contain measurement error because the theoretical concepts they represent in a model have no observable counterparts. Most economic analyses involve these latent variables. Economic concepts such as utility, ability and achievement cannot be directly measured. In other cases, data available to researchers are not the same quantities to which economic agents are assumed to react. Thus, the available data merely serve as proxy variables for the theoretical concepts being studied.

Many studies have found serious measurement error in the panel data used in common economic analyses. Bowers and Horvath (1984) found that the change in unemployment duration data between monthly surveys from the Current Population Survey (CPS) tended to exceed the actual elapsed time. They concluded that the reported length of job search contain weeks of search while a person is on the job or out of the labor force, and, thus, possibly biasing the results of studies that use this duration data. Duncan and Hill (1985) conducted a validation study of Panel Study of Income Dynamics (PSID) by comparing responses to PSID questions with highly accurate company records. They found that although only a small percentage responded incorrectly to questions soliciting categorical variables such

as union contract coverage, receipt of medical insurance and number of paid vacation days, there was strong evidence of a substantial reporting error variance for some continuous measures like annual hours worked and subsequently, other measures based on these erroneously reported responses (e.g., average hourly earnings, defined as the ratio of annual earnings to annual hours worked). Bound *et al.* (1990) conducted a similar validation study using two data sources, the PSID Validation Study (PSIDVS) and the U. S. Current Population Survey. They compared labor market measures such as annual earnings, work hours and tenure, with company records and Social Security records, respectively. In their study, the unemployment event-history data and the four-year change in the ratio of annual earnings to annual hours were found to be unreliable. Altonji and Siow (1987) also found that measurement error in income had a strong influence on the relationship between consumption and income in their tests of the rational expectations life cycle model of consumption. Similarly, Aasness, Biørn, and Skjerpen (1993) estimated a system of consumer expenditure functions using Norwegian household budget (panel) data to make inferences on Engel functions while at the same time modeling measurement errors in consumption. The hypothesis of no measurement error in total expenditure was strongly rejected in the tests they conducted.

Techniques for fitting models that account for measurement error in the explanatory variables to cross-sectional data are well-known (see Fuller, 1987). The statistical consequences of not accounting for errors in the explanatory variables include biased and inconsistent parameter estimates. When measurement error is accounted for, model parameters cannot be identified without extraneous information in the form of replication, valid instruments, or additional assumptions (Fuller, 1987). One common practice in dealing with measurement error involves the use of reliability estimates from validation studies or small studies in which the same variable is measured repeatedly from the same individual in a short period of time. However, Goldstein (1979) pointed out that getting these reliability estimates may be almost impossible in some cases. For instance, in reproducibility studies where learning can take place, measurements cannot be repeatedly taken from the same individual without allowing the measuring instrument to alter over time and dependencies between measurements to occur. Exclusion of the unobservable variable rather than using a proxy is a possible option. However, studies (McCallum, 1972; Wickens, 1972; Aigner, 1974) have found that there may still be gains to using the proxy variable.

So far, most of the existing research in panel data analysis has dealt with developing methods for models involving explanatory variables measured without error. While some studies have been devoted to the topic of measurement error in panel data, these studies are few relative to the many journal

articles and books that have discussed the analysis of panel data models without measurement error.

## Errors-in-Variables Models for Panel Data

The presence of measurement error affects the performance of standard panel data methods. Solon (1989) noted that using standard panel data methods applied to data where the measurement error is less serially correlated than the true value of the measured variable exacerbates the bias in the coefficient estimates, compared to using the usual regression methods for cross-sectional data. He cited, as an example, the estimation of wage premia for undesirable job characteristics, where the relevant model relating the wage measure for the  $i$ -th worker's job in period  $t$ ,  $Y_{it}$ , to the risk of injury on the job for the  $i$ -th worker in period  $t$ ,  $x_{it}$ , is given by

$$Y_{it} = \alpha + \beta x_{it} + \gamma_i + \epsilon_{it}, \quad i = 1, 2, \dots, n, \quad t = 1, 2, \dots, T.$$

where  $\gamma_i$ ,  $i = 1, 2, \dots, n$ , represent unknown individual-specific effects with  $E(\gamma_i) = 0$ , and  $x_{it}$  is assumed to be measured imperfectly by  $X_{it}$ ,  $X_{it} = x_{it} + u_{it}$ . Assume that  $\text{var}(x_{it}) = \sigma_x^2$ ,  $\text{cov}(x_{it}, x_{i,t-1}) = \rho_x \sigma_x^2$ ,  $\text{var}(u_{it}) = \sigma_u^2$ ,  $\text{cov}(u_{it}, u_{i,t-1}) = \rho_u \sigma_u^2$  and that for any  $t$  and  $s$ ,  $\text{cov}(x_{it}, u_{is})$ ,  $\text{cov}(\gamma_i, u_{it})$ ,  $\text{cov}(\epsilon_{it}, \epsilon_{i,t-1})$ ,  $\text{cov}(\epsilon_{it}, u_{is})$ ,  $\text{cov}(x_{it}, \epsilon_{is})$  and  $\text{cov}(\gamma_i, \epsilon_{it})$  are all zero, but  $\text{cov}(x_{it}, \gamma_i)$  is nonzero. Applying ordinary least squares (OLS) to the cross-sectional regression of  $Y_{it}$  on  $X_{it}$  in any particular period yields the estimator of  $\beta$ ,  $\hat{\beta}_{OLS}$ , with probability limit

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{OLS} = \beta + \frac{\text{cov}(x_{it}, \gamma_i)}{\sigma_x^2 + \sigma_u^2} - \frac{\beta \sigma_u^2}{\sigma_x^2 + \sigma_u^2}.$$

This estimator is inconsistent due to the nonzero correlation between  $x_{it}$  and  $\gamma_i$  and also due to the existence of measurement error. With panel data, the inconsistency due to the correlation between  $x_{it}$  and  $\gamma_i$  may be eliminated by applying OLS to the differenced equation

$$Y_{it} - Y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + (\epsilon_{it} - \epsilon_{i,t-1})$$

so that

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{diff}} = \beta - \frac{\beta \sigma_u^2}{\left(\frac{1-\rho_x}{1-\rho_u}\right) \sigma_x^2 + \sigma_u^2}.$$

Hence,  $\hat{\beta}_{\text{diff}}$  is still inconsistent since measurement error still exists. However, differencing increases the noise-to-signal ratio for the measured explanatory variable when  $\rho_x > \rho_u$ . Thus, using “changes” instead of “levels” intensifies the bias from the measurement error in the explanatory variables if  $\rho_x > \rho_u$ .

Biørn (1992) also examined the behavior, in the presence of measurement error, of standard panel data methods originally developed for models without measurement error. Only estimators based on

transforming the data to eliminate the effects of individual heterogeneity in the panel data set were studied. Most of the estimators evaluated were found to be inconsistent. However, Biørn demonstrated that, given assumptions about the distribution of the true explanatory variables, consistent estimators can be constructed by taking a weighted average of inconsistent estimators.

Griliches and Hausman (1986) studied the random effect model (1) with  $k = 1$  when  $x_{it}$  is imprecisely measured by  $X_{it}$ , where

$$X_{it} = x_{it} + u_{it}. \quad (3)$$

They suggested controlling for individual heterogeneity by taking the difference

$$Y_{it} - Y_{is} = \beta(x_{it} - x_{is}) + e_{it} - e_{is}, \quad t \neq s = 1, 2, \dots, T, \quad (4)$$

thereby eliminating  $\alpha_i$ . Furthermore, they pointed out that the model parameters can be identified without the use of external instruments. That is, various functions of  $X_{iq}$ ,  $q = 1, 2, \dots, T$ , can be used as instrumental variables for  $(x_{it} - x_{is})$  under given information about the distribution of  $x_{it}$  and  $u_{it}$ . Hence, using each equation of the form (4) and the appropriate instrumental variables for  $(x_{it} - x_{is})$ , one can obtain several initial estimators of  $\beta$ , denoted  $\hat{\beta}_j^I$ , from the  $j$ -th equation. Note, however, that only a portion of the data is used in the computation of each of these initial estimators of  $\beta$ . Thus, these initial instrumental variables estimators cannot be expected to be efficient. Moreover, one can end up with more equations (4) than parameters to estimate. That is, the model can be over-identified. As Griliches and Hausman (1986) pointed out, the initial estimators  $\hat{\beta}_j^I$  must be constrained to be equal and must be weighted appropriately for the final estimator of  $\beta$  to be asymptotically efficient. Griliches and Hausman suggested the use of the generalized method of moments method (Hansen, 1982) to combine these initial estimators.

The idea behind the generalized method of moments (GMM) is that moment conditions can be used to define model parameters as well as to test model specification (Davidson and MacKinnon, 1993). For example, suppose an estimator of  $\beta$  in the model

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\eta}$$

is sought using valid instruments  $\mathbf{W}$ , where  $\mathbf{Y}$ ,  $\mathbf{X}$  and  $\boldsymbol{\eta}$  are all  $n \times 1$  vectors, and  $\mathbf{X}$  and  $\boldsymbol{\eta}$  may be dependent on each other. The relevant orthogonality condition that can be used to derive the GMM estimator is

$$E[\mathbf{W}'\boldsymbol{\eta}] = 0.$$

Under this condition, a GMM estimator of  $\beta$  is obtained by minimizing the criterion function

$$(\mathbf{Y} - \mathbf{X}\beta)' \mathbf{W}(\mathbf{W}'\Omega\mathbf{W})^{-1} \mathbf{W}'(\mathbf{Y} - \mathbf{X}\beta), \quad (5)$$

where  $\Omega$  is an  $n \times n$  matrix. This estimator is

$$\tilde{\beta}_{\text{GMM}} = [\mathbf{X}'\mathbf{W}(\mathbf{W}'\Omega\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{W}(\mathbf{W}'\Omega\mathbf{W})^{-1} \mathbf{W}'\mathbf{Y}, \quad (6)$$

and the corresponding estimator for the variance of  $\tilde{\beta}_{\text{GMM}}$  is

$$\text{var}(\tilde{\beta}_{\text{GMM}}) = [\mathbf{X}'\mathbf{W}(\mathbf{W}'\Omega\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}]^{-1}. \quad (7)$$

If  $\Omega$  is unknown, consistent estimators of  $(\mathbf{W}'\Omega\mathbf{W})^{-1}$  may be used. As shown by White (1980), although  $\Omega$  in (7) is typically an unknown quantity, one need not seek a consistent estimator of  $\Omega$ . Rather,  $(\mathbf{W}'\Omega\mathbf{W})$  can be estimated by  $(\mathbf{W}'\hat{\Omega}\mathbf{W})$ , where  $\hat{\Omega}$  may be based on some reasonable estimator of  $\Omega$ , possibly a matrix whose  $i$ -th diagonal element is the square of the  $i$ -th residual (see also Davidson and MacKinnon, 1993, section 16.3).

Griliches and Hausman (1986) used this GMM method to combine the initial instrumental variables estimators based on a number of differences (4). For the case with  $T = 3$ , they write

$$\mathbf{Y}_1^* = \beta \mathbf{X}_1^* + \boldsymbol{\eta}_1^*,$$

$$\mathbf{Y}_2^* = \beta \mathbf{X}_2^* + \boldsymbol{\eta}_2^*,$$

$$\mathbf{Y}_3^* = \beta \mathbf{X}_3^* + \boldsymbol{\eta}_3^*,$$

where  $\mathbf{Y}_1^*$  has  $i$ -th element equal to  $(Y_{i2} - Y_{i1})$ ,  $\mathbf{Y}_2^*$  has  $i$ -th element equal to  $(Y_{i3} - Y_{i2})$ ,  $\mathbf{Y}_3^*$  has  $i$ -th element equal to  $(Y_{i3} - Y_{i1})$ ,  $i = 1, \dots, n$ , and similarly for  $\mathbf{X}_1^*$ ,  $\mathbf{X}_2^*$ , and  $\mathbf{X}_3^*$  using the observed  $X_{it}$ , not  $x_{it}$ . Thus,  $\boldsymbol{\eta}_1^*$  has  $i$ -th element equal to  $(e_{i2} - e_{i1}) - \beta(u_{i2} - u_{i1})$ ,  $\boldsymbol{\eta}_2^*$  has  $i$ -th element equal to  $(e_{i3} - e_{i2}) - \beta(u_{i3} - u_{i2})$ , and  $\boldsymbol{\eta}_3^*$  has  $i$ -th element equal to  $(e_{i3} - e_{i1}) - \beta(u_{i3} - u_{i1})$ . Initial estimates of  $\beta$  from each equation,  $\hat{\beta}_j^I$ ,  $j = 1, 2, 3$ , are obtained based on  $\mathbf{Y}_j^*$  and  $\mathbf{X}_j^*$ , using  $\mathbf{w}_{ji}$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, 3$ , as instruments. Let  $Y_{ji}^*$  and  $X_{ji}^*$  denote the  $i$ -th element of  $\mathbf{Y}_j^*$  and  $\mathbf{X}_j^*$ , respectively. Then,

Griliches and Hausman defined the combined GMM estimator of  $\beta$  according to (6) using

$$\mathbf{Y} = (Y_{11}^*, Y_{21}^*, Y_{31}^*), (Y_{12}^*, Y_{22}^*, Y_{32}^*), \dots, (Y_{1n}^*, Y_{2n}^*, Y_{3n}^*)',$$

$$\mathbf{X} = (X_{11}^*, X_{21}^*, X_{31}^*), (X_{12}^*, X_{22}^*, X_{32}^*), \dots, (X_{1n}^*, X_{2n}^*, X_{3n}^*)',$$

$$\mathbf{W} = [\bar{\mathbf{w}}_1', \bar{\mathbf{w}}_2', \dots, \bar{\mathbf{w}}_n']',$$

where

$$\bar{\mathbf{w}}_1 = \begin{bmatrix} \mathbf{w}'_{11} & 0 & 0 \\ 0 & \mathbf{w}'_{21} & 0 \\ 0 & 0 & \mathbf{w}'_{31} \end{bmatrix}, \quad \bar{\mathbf{w}}_2 = \begin{bmatrix} \mathbf{w}'_{12} & 0 & 0 \\ 0 & \mathbf{w}'_{22} & 0 \\ 0 & 0 & \mathbf{w}'_{32} \end{bmatrix}, \quad \dots, \quad \bar{\mathbf{w}}_n = \begin{bmatrix} \mathbf{w}'_{1n} & 0 & 0 \\ 0 & \mathbf{w}'_{2n} & 0 \\ 0 & 0 & \mathbf{w}'_{3n} \end{bmatrix}.$$

They suggested the use of the heteroscedastic-error weight matrix

$$(\mathbf{W}'\tilde{\Omega}\mathbf{W}) = \sum_{i=1}^n \tilde{\mathbf{w}}_i' \tilde{\epsilon}_i \tilde{\epsilon}_i' \tilde{\mathbf{w}}_i, \quad (8)$$

where

$$\begin{aligned} \tilde{\epsilon}_i &= (\tilde{\eta}_{1i}^*, \tilde{\eta}_{2i}^*, \tilde{\eta}_{3i}^*)', \quad i = 1, 2, \dots, n, \\ (\tilde{\eta}_{j1}^*, \dots, \tilde{\eta}_{jn}^*)' &= \mathbf{Y}_j^* - \beta_j^I \mathbf{X}_j^*, \quad j = 1, 2, 3. \end{aligned}$$

If both equation errors  $e_{it}$  and measurement errors  $u_{it}$  in (1) and (3) are assumed to be independent and identically distributed over  $i$  and  $t$ ,

$$\begin{aligned} \text{var} \left[ \begin{pmatrix} \eta_{1i}^* \\ \eta_{2i}^* \\ \eta_{3i}^* \end{pmatrix} \right] &= \text{var} \left[ \begin{pmatrix} (e_{i2} - e_{i1}) - \beta(u_{i2} - u_{i1}) \\ (e_{i3} - e_{i2}) - \beta(u_{i3} - u_{i2}) \\ (e_{i3} - e_{i1}) - \beta(u_{i3} - u_{i1}) \end{pmatrix} \right] \\ &= \Gamma, \quad i = 1, 2, \dots, n. \end{aligned}$$

which is common over individuals. Therefore, under the assumption that  $e_{it}$  and  $u_{it}$  are i.i.d. over  $i$  and  $t$ , more efficient estimators for  $\beta$  and  $\text{var}(\tilde{\beta}_{\text{GMM}})$ , denoted by  $\tilde{\beta}_{\text{GMM}}^*$  and  $\text{var}(\tilde{\beta}_{\text{GMM}}^*)$ , respectively, can be defined by replacing  $(\mathbf{W}'\Omega\mathbf{W})$  in (6) and (7) with

$$(\mathbf{W}'\Omega\mathbf{W}) = \sum_{i=1}^n \tilde{\mathbf{w}}_i' \tilde{\Gamma} \tilde{\mathbf{w}}_i, \quad (9)$$

where

$$\tilde{\Gamma} = \frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i \tilde{\epsilon}_i'.$$

The use of instrumental variables in estimation, however, is difficult in practice. Among the difficulties association with instrumental variables methods is verifying the validity of instruments used in the estimation. This requires information about the distribution of the true explanatory variable and the measurement error, which are both unobservable. Aside from this, the method proposed by Griliches and Hausman (1986) yields little insight about the identification of model parameters.

While there is an extensive literature on the analysis of random coefficient models like (2) (see, e.g., Swamy, 1970; Carter and Yang, 1986; Harville, 1977; Laird and Ware, 1982; Gumpertz and Pantula, 1989), no previous studies have examined the random coefficient model (2) when all or some explanatory variables are measured with error, as in (3).

## Moment Structure Analysis

Moment structure analysis has been widely used in the social and behavioral sciences for models involving latent variables. Also known in the literature as structural equation modeling, it is a flexible

and comprehensive approach that encompasses many standard statistical models, including the analysis of variance, multiple regression and factor analysis. Bollen and Long (1993) point out that structural equation models accommodate simultaneous equations with many endogenous variables and allow exogenous and endogenous variables to be measured with error. Hoyle (1995) provides a nontechnical overview of various concepts and issues in structural equation modeling. See also Bollen (1989).

### Covariance Structure Analysis

Browne (1982) gives a comprehensive review of covariance structures. The fundamental hypothesis of covariance structure analysis is that the covariance matrix of observed variables is a matrix-valued function of a set of parameters,

$$\Sigma = \Sigma(\theta).$$

The structure of  $\Sigma(\theta)$  may arise from the relationship of the observed variables to certain hypothetical unobservable or latent variables. The parameter vector  $\theta$  or the model is then said to be identified if  $\Sigma(\theta) = \Sigma(\tilde{\theta})$  implies that  $\theta = \tilde{\theta}$ . A necessary condition for  $\theta$  to be identified is that the number of elements of  $\theta$  be less than or equal to the number of non-redundant elements of  $\Sigma(\theta)$ . If the latter exceeds the number of parameters to be estimated, the model is said to be over-identified. Only over-identified models are of interest because only in this case is the issue of model fit meaningful.

If  $\theta$  is identified, an estimator  $\hat{\theta}$  is chosen by fitting the sample covariance matrix of the observed variables,  $S$ , to the covariance matrix implied by the model,  $\Sigma(\theta)$ . A scalar function which indicates the discrepancy between the sample covariance matrix  $S$  and the fitted matrix  $\Sigma$ ,  $F(S; \Sigma(\hat{\theta}))$ , called a discrepancy function, can be used to define an adequate fit. The estimator  $\hat{\theta}$  is then obtained by minimizing the discrepancy function, typically using iterative methods. The choice of discrepancy function affects the asymptotic distribution of an estimator  $\hat{\theta}$ . For the resulting estimator  $\hat{\theta}$  to be consistent,  $\Sigma(\hat{\theta})$  must be continuous, and the discrepancy function must be such that (i)  $F(S; \Sigma(\hat{\theta})) \geq 0$ , (ii)  $F(S; \Sigma(\hat{\theta})) = 0$  if and only if  $S = \Sigma(\hat{\theta})$ , and (iii)  $F(S; \Sigma(\hat{\theta}))$  is continuous in  $S$  and  $\Sigma(\hat{\theta})$  (Bollen, 1989).

The normal theory maximum likelihood function

$$F_{ML}(S; \Sigma(\theta)) = n[\ln |\Sigma(\theta)| + \text{tr}[S\Sigma^{-1}(\theta)] - \ln |S| - p], \quad (10)$$

where  $p$  is the total number of manifest or observable variables, is one of the most widely used discrepancy functions. The use of  $F_{ML}$  is based on the assumption that the observed variables have a multivariate normal distribution. Under this assumption, the resulting estimator of  $\theta$ , denoted by



$\hat{\theta}_{ML}$ , is consistent, asymptotically efficient and asymptotically normal. Using  $F_{ML}(S; \Sigma(\theta))$  in (10) as a fitting function also provides a test of overall model fit for over-identified models. If the observed variables follow the multivariate normal distribution,  $F_{ML}(S, \Sigma(\hat{\theta}_{ML}))$  is asymptotically distributed as a  $\chi^2$  random variable and can be used to test the null hypothesis  $H_0 : \Sigma = \Sigma(\theta)$  (Bollen, 1989).

Other estimators of  $\theta$  can be obtained by using other discrepancy functions. For instance, minimizing the discrepancy function

$$F_{GLS}(S; \Sigma(\theta)) = \frac{n}{2} \left[ \text{tr} \left\{ \left[ S - \Sigma(\theta) \right] W^{-1} \right\} \right] \quad (11)$$

where  $W$  is a weight matrix (typically  $S^{-1}$ ), leads to the generalized least squares estimator of  $\theta$ .

The discrepancy function (10) is often used even when the observed variables are not normally distributed. Estimation in this manner is referred to in the literature as pseudo maximum likelihood (PML). Despite the violation of the assumptions underlying the use of (10), the resulting PML estimator is still consistent. However, inferences based on PML statistics may not be valid. Browne (1984) developed asymptotic distribution-free (ADF) methods, but ADF methods tend to be computationally intensive and statistically unstable since they involve fourth-order moments. Studies have also found that very large samples are required to get the ADF-based chi-square goodness-of-fit test to perform adequately (see, e.g., Hu *et al.*, 1992 and West *et al.*, 1995). In a different line of research, PML inference procedures have been demonstrated to yield valid inferences even when the normality distributional assumption cannot be made (Browne, 1987; Anderson and Amemiya, 1988; Browne and Shapiro, 1988; Amemiya and Anderson, 1990; Papadopoulos and Amemiya, 1996). In particular, these studies have shown that for a large class of models, the  $\chi^2$  distribution is appropriate for describing the sampling distribution of goodness-of-fit tests and that normal theory PML standard errors of many important parameter estimators are correct even if normality assumptions do not hold. Thus, researchers can act as if the data are normally distributed (regardless of whether or not they really are normally distributed), obtain parameter estimates via maximum normal likelihood and still make valid inferences about some parameters. These findings suggest that moment structure analysis has a wide degree of applicability.

### Mean and Covariance Structure Analysis

Unlike traditional structural equation models in which means are usually left unrestricted, some applications require imposing a structure not only on the covariance matrix of observed variables but also on the mean of the observed variables. Then the model consists of  $\mu(\theta)$  and  $\Sigma(\theta)$ , where  $\theta$  is some vector of parameters. Under the assumption of normal observations, the maximum likelihood estimator

of  $\theta$  is obtained by minimizing

$$F_{ML}^*(\bar{\mathbf{Z}}, \mathbf{S}; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)) = n \{ [\bar{\mathbf{Z}} - \boldsymbol{\mu}(\theta)]' \boldsymbol{\Sigma} [\bar{\mathbf{Z}} - \boldsymbol{\mu}(\theta)] + \ln |\boldsymbol{\Sigma}(\theta)| + \text{tr}[\mathbf{S} \boldsymbol{\Sigma}^{-1}(\theta)] - \ln |\mathbf{S}| - p \} . \quad (12)$$

where  $\bar{\mathbf{Z}}$  and  $\mathbf{S}$  are the sample mean vector and covariance matrix, respectively (Browne and Arminger, 1995).

Instead of having to modify the fitting function used in conventional software for covariance structure analysis, an alternative approach to mean and covariance structure analysis is to analyze the uncorrected second-order moment structure of the vector of observable variables augmented with a constant equal to one (Satorra, 1992). That is, packages that minimize (10) can be used to obtain  $\hat{\theta}_{ML}$  by substituting the augmented moment matrices

$$\mathbf{S}^* = \begin{bmatrix} \mathbf{S} + \bar{\mathbf{Z}}\bar{\mathbf{Z}}' & \bar{\mathbf{Z}}' \\ \bar{\mathbf{Z}} & 1 \end{bmatrix}$$

and

$$\boldsymbol{\Omega}^*(\theta) = \begin{bmatrix} \boldsymbol{\Omega}(\theta) + \boldsymbol{\mu}(\theta)\boldsymbol{\mu}(\theta)' & \boldsymbol{\mu}(\theta)' \\ \boldsymbol{\mu}(\theta) & 1 \end{bmatrix}$$

for  $\mathbf{S}$  and  $\boldsymbol{\Omega}(\theta)$  in (10) to yield the same estimator that would have been obtained by minimizing (12). The error degrees of freedom given by the software needs to be decreased by one to obtain the correct value.

## Dissertation Organization

This dissertation consists of two papers on linear errors-in-variables models for data consisting of continuous responses measured at several time points. The models are static in nature; that is, no lagged dependent variables are included among the explanatory variables.

The first paper considers identification and estimation of a random effect errors-in-variables model for panel data. Here, individual heterogeneity is assumed to be manifested in intercepts that randomly differ around a common mean across individuals. This random effect is also assumed to represent omitted individual-specific but time-independent characteristics not accounted for by the explanatory variables (Mundlak, 1978). The existing estimation method (Griliches and Hausman, 1986) utilizes instrumental variables that can be obtained from within the panel data set. The paper proposes a new and more general approach to the identification and estimation of the model. This approach is the essentially moment structure analysis approach. Numerical results that compare the performance of the proposed method with the existing method are reported.

In the second paper, a generalization of the random effect errors-in-variables model that allows the slope coefficients to differ across individuals is studied. This random coefficient errors-in-variables model has not been studied before. Consequently, no estimation method for this model exists. The paper examines the identification issue for this random coefficient measurement error model. In addition, methods of estimating the parameters of the model are proposed. Simulation results showing the finite sample behavior of the proposed estimation methods are presented.

The dissertation concludes with a summary of the results of these two studies as well as a discussion on possible areas for further research. Additional simulation results for the random effect errors-in-variables model are given in Appendix 1 while Appendix 2 presents additional simulation results for the random coefficient errors-in-variables model.

## References

- Aasness, K., E. Biørn and T. Skjerpen, 1993, Engel functions, panel data, and latent variables, *Econometrica* 61, 1395-1422.
- Aigner, D. J., 1974, MSE dominance of least squares with errors-of-observation, *Journal of Econometrics* 2, 365-372.
- Altonji, J. G. and A. Siow, 1987, Testing the response of consumption to income changes with (noisy) panel data, *Quarterly Journal of Economics* 102, 293-328.
- Amemiya, Y. and T. W. Anderson, 1990, Asymptotic chi-square tests for a large class of factor analysis models, *Annals of Statistics* 18, 1453-1463.
- Anderson, T. W. and Y. Amemiya, 1988, The asymptotic normal distribution of estimators in factor analysis under general conditions, *Annals of Statistics* 16, 757-771.
- Baltagi, B. H., 1995a, Panel data, *Journal of Econometrics* 68 (North-Holland, Amsterdam).
- Baltagi, B. H., 1995b, *Econometric Analysis of Panel Data* (John Wiley, New York).
- Baltagi, B. H. and B. Raj, 1992, A survey of recent theoretical developments in the econometrics of panel data, *Empirical Economics* 17, 85-109.
- Biørn, E., 1992, Panel data with measurement errors, in: L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of Theory and Applications* (Kluwer Academic Publishers, Dordrecht) 152-195.
- Bollen, K. A., 1989, *Structural Equations with Latent Variables* (John Wiley, New York).
- Bollen, K. A. and J. S. Long, 1993, *Testing Structural Equation Models* (Sage Publications, Thousand Oaks).

- Bound, J., C. Brown, G. J. Duncan and W. L. Rogers. 1990. Measurement error in cross-sectional and longitudinal labor market surveys: validation study evidence. in: J. Hartog, G. Ridder and T. Theeuwes, eds., *Panel Data and Labor Market Studies* (Elsevier Science Pub. Co., New York).
- Bowers, N. and F. W. Horvath, 1984, Keeping time: an analysis of errors in the measurement of unemployment duration, *Journal of Business and Economic Statistics* 2, 140-149.
- Browne, M. W., 1982, Covariance structures, in: D. M. Hawkins, ed., *Topics in Applied Multivariate Analysis* (Cambridge University Press, New York).
- Browne, M. W., 1984, Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology* 37, 62-83.
- Browne, M. W., 1987, Robustness of statistical inference in factor analysis and related models, *Biometrika* 74, 375-384.
- Browne, M. W. and G. Arminger, 1995, Specification and estimation of mean and covariance structure models, in: G. Arminger, C. C. Clogg, M. E. Sobel, eds., *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (Plenum Press, New York).
- Browne, M. W. and A. Shapiro, 1988, Robustness of normal theory methods in the analysis of linear latent variate models, *British Journal of Mathematical and Statistical Psychology* 41, 193-208.
- Carter, R. L. and M. C. K. Yang, 1986, Large sample inference in random coefficient regression models, *Communications in Statistics Part A-Theory and Methods* 15, 2507-2525.
- Chamberlain, G., 1984, Panel data, in: Z. Griliches and M. P. Intriligator, eds., *Handbook of Econometrics* Volume II (North-Holland, Amsterdam).
- Crowder, D. and D. Hand, 1990, *Analysis of Repeated Measures* (Chapman and Hall, London).
- Davidson, R. and J. G. MacKinnon, 1993, *Estimation and Inference in Econometrics* (Oxford University Press, New York).
- Duncan, G. J. and D. H. Hill, 1985, An investigation of the extent and consequences of measurement error in labor economic survey data, *Journal of Labor Economics* 3, 508-532.
- Fuller, W. A., 1987, *Measurement Error Models* (John Wiley, New York).
- Goldstein, H., 1979, Some models for analysing longitudinal data on educational attainment, *Journal of Royal Statistical Society Series A* 142, 407-32.
- Greene, W. H., 1993, *Econometric analysis* (Macmillan, New York).
- Griliches, Z. and J. A. Hausman, 1986, Errors in variables in panel data, *Journal of Econometrics* 31, 93-118.
- Gumpertz, M. and S. G. Pantula, 1989, A simple approach to inference in random coefficient models, *The American Statistician* 43, 203-210.
- Hansen, L., 1982, Large sample properties of generalized method of moments, *Econometrica* 50, 1029-54.

- Harville, D. A., 1977, Maximum likelihood approaches to variance components estimation and to related problems, *Journal of the American Statistical Association* 72, 320-340.
- Heckman, J. J. and B. Singer, 1982, Econometrics analysis of longitudinal data, *Journal of Econometrics* 18, 1-169.
- Hoyle, R., 1995, *Structural Equation Models: Concepts, Issues and Applications* (Sage Publications, Thousand Oaks).
- Hsiao, C., 1986, *Analysis of Panel Data* (Cambridge University Press, New York).
- Hu, L., P.M. Bentler, and Y. Kano, 1992, Can test statistics in covariance structure analysis be trusted?, *Psychological Bulletin* 112, 351-362.
- Johnston, J., 1984, *Econometric Methods* 3rd ed. (McGraw-Hill, New York).
- Judge, G. G., R. C. Hill, W. E. Griliches, H. Lütkepohl and T. C. Lee., 1985, *The Theory and Practice of Econometrics* 2nd ed. (John Wiley, New York).
- Laird, N. M. and J. H. Ware, 1982, Random-effects models for longitudinal data, *Biometrics* 38, 963-974.
- Maddala, G. S., 1987, Recent developments in the econometrics of panel data, *Transportation Research* 21, 303-338.
- Matyas, L. and P. Sevestre, 1992, *The Econometrics of Panel Data: Handbook of Theory and Applications* (Kluwer Academic Publishers, Dordrecht).
- McCallum, B., 1972, Relative asymptotic bias from errors of omission and measurement, *Econometrica* 40, 757-758.
- Morgenstern, O., 1963, *On the Accuracy of Economic Observations* (Princeton University Press, New Jersey).
- Mundlak, Y., 1978, On the pooling of cross-section and time-series data, *Econometrica* 46, 69-85.
- Papadopoulos, S. and Y. Amemiya, 1996, On linear latent variable analysis of multiple populations, Preprint Series No. 96-31, Statistical Laboratory, Iowa State University.
- Satorra, A., 1992, Asymptotic robust inferences in the analysis of mean and covariance structures, in: P.V. Marsden *Sociological Methodology*.
- Solon, G. S., 1989, The value of panel data in economic research, in: D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh, eds., *Panel Surveys* (John Wiley, New York).
- Swamy, P. A. V. B., 1970, Efficient inference in random coefficient regression models, *Econometrica* 38, 311-323.
- West, S. G., J. F. Finch and P. J. Curran, 1995, Structural equation models with nonnormal data, in: R. Hoyle, ed., *Structural Equation Models: Concepts, Issues and Applications* (Sage Publications, Thousand Oaks).

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817-838.

Wickens, M. R., 1972, A note on the use of proxy variables, *Econometrica* 40, 759-61.

## ANALYSIS OF PANEL DATA USING A RANDOM EFFECT ERRORS-IN-VARIABLES MODEL

A paper to be submitted to the Journal of Econometrics

Elizabeth Martha Paterno and Yasuo Amemiya

### Abstract

Random effect analysis for panel data is considered when some explanatory variables are measured with error. In some applications (e.g., economic analysis), the covariance between the random effect and the unobservable true explanatory variables is to be estimated and contributes to the difficulty of the problem. Identification of model parameters given the first two moments of observed variables is examined, and relatively unrestrictive sufficient conditions for identification are obtained. Estimation based on maximum normal likelihood is proposed. This method can be easily implemented using available computer packages that perform moment structure analysis. Compared to the only existing procedure based on instrumental variables, the new method is shown to be more efficient and to have much wider applicability. Standard error estimates and goodness-of-fit statistics obtained under the assumption of normally distributed observations are shown to be asymptotically valid for a broad class of non-normal observations. Simulation results demonstrating the efficiency and usefulness of the new procedure are presented.

### Introduction

Panel data sets consist of observations from a number of individuals repeatedly measured over time. Alternatively referred to as longitudinal data, they arise often in various fields. In many applications in economics and other social sciences, time points in which measurements are taken are equally spaced and are usually common for all individuals. Moreover, the number of time points tends to be small relative to the number of units from which measurements are taken. Thus, when a dependent response

variable  $Y$  and a  $k \times 1$  covariate vector  $\mathbf{x}$  are measured. a typical panel data set consists of  $(Y_{it}, \mathbf{x}'_{it})$ ,  $i = 1, 2, \dots, n$ ,  $t = 1, 2, \dots, T$ , where  $n$  is the number of individuals, and  $T$  is the number of time points common to all individuals. Here,  $\mathbf{x}_{it}$ , in general, represents a time- and individual-dependent covariate vector. Using these data, researchers are often interested in fitting regression models of the type

$$Y_{it} = \alpha_0 + \beta'_0 \mathbf{x}_{it} + e_{it}^0,$$

where the  $e_{it}^0$ 's represent random errors with mean zero.

Observations taken from an individual over time are usually expected to contain some nontrivial individual-specific characteristic. To account for this individual variability, statistical models containing individual random components are used in panel data regression analysis. One such model, the random effect model, assumes that

$$\begin{aligned} Y_{it} &= \alpha_i + \beta' \mathbf{x}_{it} + e_{it}, \\ \alpha_i &\sim (\mu_\alpha, \sigma_{\alpha\alpha}), \\ e_{it} &\sim (0, \sigma_{ee}), \end{aligned} \tag{1}$$

where  $\alpha_i$  and  $e_{it}$  are independent, and  $\alpha_i$ ,  $i = 1, 2, \dots, n$ , represent random unobserved individual effects possibly correlated with  $\mathbf{x}_{it}$ . For this population of individuals, the average intercept is  $\mu_\alpha = E[\alpha_i]$ , and  $\sigma_{\alpha\alpha} = V[\alpha_i]$  is the variability in the response  $Y_{it}$  due to the individual differences.

However, the  $k \times 1$  vector of regressors  $\mathbf{x}_{it}$  often cannot be precisely measured. The existence of measurement errors in most economic data has long been recognized. In fact, many studies have found serious measurement error in the panel data used in common economic analyses (see, e.g., Bowers and Horvath, 1984; Duncan and Hill, 1985; Altonji and Siow, 1987; Bound *et al.*, 1990; Aasness, Biørn, and Skjerpen, 1993). Not only do measurement errors arise from the survey process in the form of transcription, transmission or recording errors, variables may also be said to contain measurement error because the theoretical concepts they represent have no observable counterparts. For instance, concepts in economics such as utility, ability and achievement cannot be directly measured. In addition, data available to researchers may not be the same quantities to which economic agents react. Thus, researchers have to use the observed but fallible measure,  $\mathbf{X}_{it}$ , where

$$\mathbf{X}_{it} = \mathbf{x}_{it} + \mathbf{u}_{it}. \tag{2}$$

Here, the measurement error  $\mathbf{u}_{it}$  is assumed to have mean zero and to be independent of  $\mathbf{x}_{it}$  and  $e_{it}$ . The elements of  $\mathbf{u}_{it}$  corresponding to error-free variables are zero, but the problem is formulated here as



if all elements of  $\mathbf{x}_{it}$  are measured with error. In most problems in economics, the measurement errors for different components of  $\mathbf{x}_{it}$  can be treated as independent, but errors may be correlated over time.

The statistical consequences of not accounting for errors in explanatory variables in models for cross-sectional data include biased and inconsistent parameter estimators. When measurement error is accounted for, model parameters cannot be identified without extraneous information in the form of replication, valid instruments, or additional assumptions (Fuller, 1987). The presence of measurement error also affects the performance of standard panel data methods. Solon (1989, p. 494) noted that "estimation procedures commonly used with panel data are especially vulnerable to bias from measurement error if the measurement error is less serially correlated than is the true value of the measured variable." While many studies have examined measurement error models using cross-sectional data, the random effect model (1) with measurement error structure (2) has not been discussed widely, but is clearly an important problem in economics and other related fields.

The usual statistical formulation of model (1) does not allow interaction of  $\alpha_i$  and the individual explanatory variables. However, in some economic applications, researchers assume nonzero correlation between the random intercept  $\alpha_i$  and the covariate  $\mathbf{x}_{it}$ . This idea is based on the assumption that the random intercept  $\alpha_i$  contains the effect of time-independent, individual-specific covariates that are either not included in the equation or not observable (Mundlak, 1978). Not accounting for such a correlation between  $\alpha_i$  and the covariate  $\mathbf{x}_{it}$  leads to biased estimates of  $\beta$  in (1). Hence, estimating the parameters of the model assuming the possible existence of such correlations and making inferences about the correlations may be desirable.

Upon combining this correlation issue and the possible correlation of the elements of  $(\mathbf{u}_{it}, e_{it})$  over time, one can express the random effect model (1) with errors-in-variables as follows. For  $i = 1, 2, \dots, n$ , let

$$\begin{aligned} \mathbf{Y}_i &= (Y_{i1}, Y_{i2}, \dots, Y_{iT})', \\ \mathbf{e}_i &= (e_{i1}, e_{i2}, \dots, e_{iT})', \\ \mathbf{X}_i &= (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{iT})', \\ \mathbf{x}_i &= (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})', \\ \mathbf{u}_i &= (\mathbf{u}_{i1}, \mathbf{u}_{i2}, \dots, \mathbf{u}_{iT})'. \end{aligned}$$

Then, model (1)-(2) with associated assumptions can be written as

$$\mathbf{Y}_i = \alpha_i \mathbf{1}_T + \mathbf{x}_i \beta + \mathbf{e}_i, \quad (3a)$$

$$\mathbf{X}_i = \mathbf{x}_i + \mathbf{u}_i, \quad (3b)$$

where  $\alpha_i$ ,  $e_i$  and  $u_i$  are independent of each other and identically distributed over individuals  $i = 1, 2, \dots, n$ , and  $x_i$  is independent of  $e_i$  and  $u_i$ . If  $x_i$ 's are also treated as identically distributed  $T \times k$  random matrices, then we have

$$E \left[ \begin{pmatrix} \alpha_i \\ \text{vec } x_i \end{pmatrix} \right] = \begin{pmatrix} \mu_\alpha \\ \mu_x \end{pmatrix}, \quad (3c)$$

$$V \left[ \begin{pmatrix} \alpha_i \\ \text{vec } x_i \end{pmatrix} \right] = \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma'_{x\alpha} \\ \sigma_{x\alpha} & \Sigma_{xx} \end{bmatrix}. \quad (3d)$$

$$E[e_i] = 0, \quad E[u_i] = 0, \quad (3e)$$

$$V[e_i] = \Sigma_{ee}. \quad (3f)$$

$$V[\text{vec } u_i] = \Sigma_{uu} = \begin{bmatrix} \Sigma_{uu_1} & 0 & \dots & 0 \\ 0 & \Sigma_{uu_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{uu_k} \end{bmatrix}, \quad (3g)$$

where the vec operator stacks the columns of a matrix in a single vector starting with the first column.  $\Sigma_{uu}$ , is  $T \times T$ ,  $\mu_x$  is  $Tk \times 1$ ,  $\sigma_{x\alpha}$  is  $Tk \times 1$  and  $\Sigma_{xx}$  is  $Tk \times Tk$ . Although the random sample assumption for  $x_i$  will be dropped later, the form of the first two moments in (3c)-(3g) is useful in discussing identification and estimation issues.

A limited literature exists for estimation of special cases of model (3). Griliches and Hausman (1986) studied model (3) with  $k = 1$ . They suggested eliminating  $\alpha_i$  in (3a)-(3b) by taking the difference

$$Y_{it} - Y_{is} = \beta(x_{it} - x_{is}) + \epsilon_{it} - \epsilon_{is}, \quad t \neq s = 1, 2, \dots, T, \quad (4)$$

and estimating  $\beta$  using (4). Furthermore, they pointed out that the model parameters can be identified without the use of external instruments. That is, various functions of  $X_{iq}$ ,  $q = 1, 2, \dots, T$ , can be used as instrumental variables for  $(x_{it} - x_{is})$  given information about the distribution of  $x_i$  and  $u_i$ . By forming all possible equations of the form (4) and using the appropriate instruments for  $(x_{it} - x_{is})$ , several (initial) estimators of  $\beta$  can be obtained, one estimator per equation. Griliches and Hausman suggested the use of the generalized method of moments (GMM) method (Hansen, 1982) to combine these instrumental variables estimators of  $\beta$ . However, this IV/GMM method is difficult to use in practice. Finding and verifying the validity of instrumental variables used is difficult, with the degree of difficulty increasing as more variables measured with error are included in the model. The use of instrumental variables also requires that assumptions be made about the distribution of  $x_i$  and  $u_i$ . Conditions required for the validity of a particular choice of the instrumental variables involve these

assumptions and are very specific, and their violation leads to the use of invalid instruments. Even with a valid set of instruments, this estimation method may not be the most efficient. Furthermore, only  $\beta$  is estimated, and no other parameter of the original model is estimated. Lastly, this approach does not provide useful insights into the overall identification problem.

Wansbeek and Koning (1989) also considered model (3) with  $k = 1$ ,  $\sigma_{\varepsilon\alpha} = 0$ , and independently and identically distributed measurement errors,  $u_{it}$ . For this simple model, they suggested the use of maximum likelihood estimation under the normality of all variables, and they discussed the identification issue. They argued that this approach may be preferable to the GMM approach under the assumption of normality since the fourth-order moments used in GMM estimation tend to be more volatile compared to second-order moments. They derived local identification conditions for  $\beta$  based on using the (corrected) covariance matrix of the observed variables and found that a sufficient condition for local identification is that the  $\Sigma_{xx}$  must have different diagonal elements or different off-diagonal elements. These identification results are incomplete in the sense that information in the sample mean vector is ignored. Hsiao and Taylor (1991) also discussed the identification of model (3) but with fixed effects  $\alpha_i$ . Biørn (1992) examined model (3) with  $k = 1$  and focused on estimation procedures that eliminate the individual effect  $\alpha_i$ . He demonstrated that, given certain assumptions about the distribution of  $\mathbf{x}_i$ , consistent estimators can be constructed by taking a weighted average of inconsistent estimators.

## General Approach and Identification

In model (3), let

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_i \\ \text{vec } \mathbf{X}_i \end{pmatrix} \quad (5)$$

be the  $T(k+1) \times 1$  vector of observations from the  $i$ -th individual. Define

$$E[\mathbf{Z}_i] = \boldsymbol{\gamma}(\boldsymbol{\theta}), \quad (6a)$$

$$V[\mathbf{Z}_i] = \boldsymbol{\Omega}(\boldsymbol{\theta}), \quad (6b)$$

where  $\boldsymbol{\theta}$  is a vector of all unknown parameters in (3). Note that model (3) does not specify a particular distributional form of the random variables involved except for the first two moments. Thus, it seems sensible to discuss identification of the model parameter  $\boldsymbol{\theta}$  only through the first two moments  $\boldsymbol{\gamma}(\boldsymbol{\theta})$  and  $\boldsymbol{\Omega}(\boldsymbol{\theta})$ . That is, we investigate whether  $\boldsymbol{\theta}$  can be uniquely determined given  $\boldsymbol{\gamma}(\boldsymbol{\theta})$  and  $\boldsymbol{\Omega}(\boldsymbol{\theta})$ . If all variables in the model are normally distributed, this approach corresponds to the identification of the whole model. Without a particular distributional assumption (and possibly with non-random or

dependent  $\mathbf{x}_i$ ), the approach based on the first two moments provides a way to discuss the identification of the model parameter  $\theta$  without relying on a special distributional structure such as nonzero third moments. This approach falls under the general method known as structural equation modeling or moment structure analysis that is widely used in many social and behavioral sciences. See, e.g., Bollen (1989), Bollen and Long (1993) and Hoyle (1995).

For model (3), the moment structures (6) are given by

$$\gamma(\theta) = \begin{pmatrix} \gamma_Y \\ \gamma_X \end{pmatrix} = \begin{pmatrix} \mu_\alpha \mathbf{1}_T + (\beta' \otimes \mathbf{I}_T) \mu_x \\ \mu_x \end{pmatrix}. \quad (7)$$

$$\begin{aligned} \Omega(\theta) &= \begin{bmatrix} \Omega_{YY} & \Omega'_{XY} \\ \Omega_{XY} & \Omega_{XX} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1}_T & (\beta' \otimes \mathbf{I}_T) \\ 0 & \mathbf{I}_{kT} \end{bmatrix} \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma'_{x\alpha} \\ \sigma_{x\alpha} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \mathbf{1}'_T & 0 \\ (\beta \otimes \mathbf{I}_T) & \mathbf{I}_{kT} \end{bmatrix} + \begin{bmatrix} \Sigma_{ee} & 0 \\ 0 & \Sigma_{uu} \end{bmatrix}, \end{aligned} \quad (8)$$

with  $\theta = (\beta', \mu_\alpha, \mu'_x, (\text{vech } \Sigma_{ee})', (\text{vech } \Sigma_{uu_1})', \dots, (\text{vech } \Sigma_{uu_k})', \sigma_{\alpha\alpha}, \sigma'_{x\alpha}, (\text{vech } \Sigma_{xx})')'$ , where  $\text{vech } \mathbf{A}$  denotes the vector created by listing the elements of matrix  $\mathbf{A}$  on and below its diagonal starting with the first column. Therefore, the parameters of model (3) are to be determined using the following equations:

$$\gamma_Y = \mu_\alpha \mathbf{1}_T + (\beta' \otimes \mathbf{I}_T) \mu_x, \quad (9a)$$

$$\gamma_X = \mu_x, \quad (9b)$$

$$\Omega_{YY} = \sigma_{\alpha\alpha} \mathbf{1}_T \mathbf{1}'_T + (\beta' \otimes \mathbf{I}_T) \Sigma_{xx} (\beta \otimes \mathbf{I}_T) + \mathbf{1}_T \sigma'_{x\alpha} (\beta \otimes \mathbf{I}_T) + (\beta' \otimes \mathbf{I}_T) \sigma_{x\alpha} \mathbf{1}'_T + \Sigma_{ee}, \quad (9c)$$

$$\Omega_{XY} = \sigma_{x\alpha} \mathbf{1}'_T + \Sigma_{xx} (\beta \otimes \mathbf{I}_T), \quad (9d)$$

$$\Omega_{XX} = \Sigma_{xx} + \text{blockdiag}(\Sigma_{uu_1}, \dots, \Sigma_{uu_k}). \quad (9e)$$

For the parameters important in practice, namely the average intercept  $\mu_\alpha$  and the regression coefficient  $\beta$ , the following is a sufficient condition for identification.

**Theorem 1** *The parameters  $\mu_\alpha$ ,  $\mu_x$  and  $\beta$  are identified if*

$$\text{rank} \begin{bmatrix} \mathbf{1}_T & \mathbf{M}_x \end{bmatrix} = k + 1, \quad (10)$$

where  $\mathbf{M}_x = E[\mathbf{x}_i]$ .

**Proof:** Clearly,  $\mu_x$  is identified from (9b). Given  $\mu_x$  from (9b),  $\mu_\alpha$  and  $\beta$  are uniquely determined by (9a) under (10).  $\square$

Hence,  $\mu_\alpha$ ,  $\mu_x$  and  $\beta$  are identified if  $T$  is larger than  $k$  and if no redundancy in the columns of  $\mathbf{x}_i$  is expected on average. This implies that the population mean of any of the explanatory variables cannot be constant over time. Also, only information from the mean of the observed variables is needed to identify  $\mu_\alpha$  and  $\beta$ , provided condition (10) holds. Note that there is no restriction on the form of  $\sigma_{x\alpha}$ .  $\Sigma_{xx}$ ,  $\Sigma_{ee}$  and the  $T \times T$  measurement error covariance matrices  $\Sigma_{uu_j}$ ,  $j = 1, 2, \dots, k$ . For the case where  $\mathbf{x}_i$ 's are not considered a random sample, the condition in (10) with  $\mathbf{M}_x = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  is a sufficient identification condition for  $\mu_\alpha$ ,  $\mu_x$  and  $\beta$  in the sense of the existence of consistent estimators.

In checking the identification of the other parameters in (3), we introduce a large class of models for the  $T \times 1$  error vectors  $\mathbf{e}_i$  and  $\mathbf{u}_i^{(j)} = (u_{i1}^{(j)}, \dots, u_{iT}^{(j)})$ ,  $j = 1, 2, \dots, k$ . Any time series, heteroscedasticity, or other types of models express the  $T \times T$  covariance matrix in terms of less than  $\frac{T(T+1)}{2}$  parameters. Virtually all practical models have a covariance matrix belonging to the following class:

**Definition** A parametric model  $\Sigma(\gamma)$  for a  $T \times T$  covariance matrix is said to be **identified through contrasts** if  $\gamma$  can be uniquely determined from  $\Sigma(\gamma)\mathbf{Q}$  for some  $T \times (T-1)$  matrix  $\mathbf{Q}$  satisfying  $\mathbf{1}_T' \mathbf{Q} = 0$ .

The  $T \times T$  covariance matrix of a  $T \times 1$  random vector with essentially any time-dependent structure (except the unrestricted model) satisfies this definition. For example, the heteroscedastic uncorrelated (diagonal) structure and any AR or MA model give a covariance matrix satisfying this definition. For the identification of model (3), we consider two cases depending on whether  $\sigma_{x\alpha}$  is assumed to be zero. Note that the assumption that  $\sigma_{x\alpha} \neq 0$  is sometimes made, while the usual statistical formulation of the random effect model specifies that  $\sigma_{x\alpha} = 0$ .

**Theorem 2** Assume that  $\sigma_{x\alpha}$  is known to be zero, and that (10) holds. Assume also that  $\beta_j \neq 0$ ,  $j = 1, \dots, k$ , and  $\Sigma_{ee}$  is identified through contrasts. Then, all parameters in model (3) are identified.

**Proof:** By Theorem 1, under (10),  $\mu_\alpha$ ,  $\mu_x$  and  $\beta$  are identified from the mean vector. For the covariance matrix components (9c)-(9e) with  $\sigma_{x\alpha} = 0$ , we consider a transformation given by

$$\begin{bmatrix} \mathbf{I}_T & -(\beta' \otimes \mathbf{I}_T) \\ 0 & \mathbf{I}_{kT} \end{bmatrix} \begin{bmatrix} \Omega_{YY} & \Omega'_{XY} \\ \Omega_{XY} & \Omega_{XX} \end{bmatrix} \begin{bmatrix} \mathbf{I}_T & 0 \\ -(\beta \otimes \mathbf{I}_T) & \mathbf{I}_{kT} \end{bmatrix} = \begin{bmatrix} \Omega_{YY}^* & \Omega_{XY}^{*'} \\ \Omega_{XY}^* & \Omega_{XX}^* \end{bmatrix}. \quad (11)$$

Here,

$$\begin{aligned} \Omega_{YY}^* &= \sigma_{\alpha\alpha} \mathbf{1}_T \mathbf{1}_T' + (\beta' \otimes \mathbf{I}_T) \Sigma_{uu} (\beta \otimes \mathbf{I}_T) + \Sigma_{ee}, \\ \Omega_{XY}^* &= -\text{blockdiag}(\Sigma_{uu_1}, \dots, \Sigma_{uu_k}) (\beta \otimes \mathbf{I}_T) \end{aligned} \quad (12a)$$

$$= \begin{bmatrix} -\beta_1 \Sigma_{uu_1} \\ \vdots \\ -\beta_k \Sigma_{uu_k} \end{bmatrix}. \quad (12b)$$

$$\Omega_{XX}^* = \Sigma_{xx} + \text{blockdiag}(\Sigma_{uu_1}, \dots, \Sigma_{uu_k}). \quad (12c)$$

Given  $\beta_j \neq 0$ ,  $\Sigma_{uu_j}$ ,  $j = 1, 2, \dots, k$ , can be determined using (12b). Using these values of  $\Sigma_{uu_j}$ ,  $\Sigma_{xx}$  can be solved for in (12c). Next, using  $Q$  for the contrast-identified  $\Sigma_{ee}$ , transform (12a) to get

$$\begin{aligned} \Omega_{YY}^* &= \Omega_{YY}^* \begin{bmatrix} \mathbf{1}_T & Q \end{bmatrix} \\ &= \begin{bmatrix} T\sigma_{\alpha\alpha}\mathbf{1}_T & \mathbf{0}_{T \times (T-1)} \end{bmatrix} + \begin{bmatrix} \Sigma_{ee}\mathbf{1}_T & \Sigma_{ee}Q \end{bmatrix} \\ &\quad + (\beta' \otimes \mathbf{I}_T) \Sigma_{uu} (\beta \otimes \mathbf{I}_T) \begin{bmatrix} \mathbf{1}_T & Q \end{bmatrix}. \end{aligned} \quad (13)$$

Then, given the last  $(T-1)$  columns of  $\Omega_{YY}^*$  as well as  $\beta$  and  $\Sigma_{uu_j}$ ,  $j = 1, \dots, k$ ,  $\Sigma_{ee}$  is identified since all unknown elements in  $\Sigma_{ee}$  can be determined from  $\Sigma_{ee}Q$ . Once  $\Sigma_{ee}$  is solved for, (13) also gives  $\sigma_{\alpha\alpha}$ .  $\square$

Thus, in the usual statistical formulation with  $\sigma_{\alpha\alpha} = 0$ , model (3) is identified with unrestricted  $\Sigma_{uu_j}$ . Of course, the identification result still holds if any time series or any other model for  $u_i^{(j)}$  is assumed. In addition, equation errors  $e_{it}$  need not be assumed independent and identically distributed and can have almost any type of time series models. In (12b), if  $\beta_j = 0$ , then only the corresponding  $\Sigma_{uu_j}$  is not identified. Thus, in a model with some explanatory variables measured without error, the coefficients of the error-free variables may be zero and the whole model is still identified. In practice, the condition of nonzero  $\beta_j$  is not restrictive. Note that even with  $\beta_j = 0$ , the whole  $\beta$  is identified. Therefore, it is possible to test the condition  $\beta_j = 0$ . If  $\beta_j$  is determined to be zero, the corresponding  $x$ -variable can be dropped from the model, and all parameters in the reduced model are identified.

If  $\sigma_{\alpha\alpha}$  cannot be assumed to be zero and is an unknown parameter to be estimated, some restrictions on the distribution of  $u_i^{(j)}$ ,  $j = 1, 2, \dots, k$ , are required for identification. The next result gives a simple set of fairly unrestrictive sufficient conditions that apply in this case.

**Theorem 3** *Assume that (10) holds and that each of  $\Sigma_{ee}$ ,  $\Sigma_{uu_j}$ ,  $j = 1, 2, \dots, k$ , is identified through contrasts. Assume also that  $\beta_j \neq 0$ ,  $j = 1, 2, \dots, k$ . Then, all parameters including  $\sigma_{\alpha\alpha}$  in model (3) are identified.*

**Proof:** As in the proof for Theorem 2, consider a transformation of  $\Omega(\theta)$  given by (11) so that (12a)

and (12c) hold, and (12b) becomes

$$\Omega_{XY}^* = \begin{pmatrix} \Omega_{X_1Y}^* \\ \Omega_{X_2Y}^* \\ \vdots \\ \Omega_{X_kY}^* \end{pmatrix} = \sigma_{x\alpha} \mathbf{1}'_T - \begin{bmatrix} \beta_1 \Sigma_{uu_1} \\ \vdots \\ \beta_k \Sigma_{uu_k} \end{bmatrix}. \quad (14)$$

Then, transform each  $T \times T$  sub-matrix  $\Omega_{X,Y}^*$  in (14) using

$$\Omega_{X,Y}^{**} = \Omega_{X,Y}^* [ \mathbf{1}_T \quad \mathbf{Q}_j^u ],$$

where  $\mathbf{Q}_j^u$  corresponds to the contrast-identified  $\Sigma_{uu_j}$ . This yields

$$\Omega_{XY}^{**} = \begin{bmatrix} T\sigma_{x\alpha} & 0_{Tk \times (T-1)} \end{bmatrix} - \begin{bmatrix} \beta_1 \Sigma_{uu_1} \mathbf{1}_T & \beta_1 \Sigma_{uu_1} \mathbf{Q}_1^u \\ \vdots & \vdots \\ \beta_k \Sigma_{uu_k} \mathbf{1}_T & \beta_k \Sigma_{uu_k} \mathbf{Q}_k^u \end{bmatrix}. \quad (15)$$

Thus,  $\Sigma_{uu_j}$ ,  $j = 1, 2, \dots, k$ , are identified from the last  $(T-1)$  columns of  $\Omega_{XY}^{**}$  using  $\beta_j \neq 0$  and the contrast identification condition. Once the  $\Sigma_{uu_j}$ 's are determined, (15) also yields  $\sigma_{x\alpha}$ . The argument used for (13) proves the identification of  $\Sigma_{ee}$  and  $\sigma_{\alpha\alpha}$ .  $\square$

Therefore, for a large class of practical models for  $e_{it}$  and  $u_i^{(j)}$ , the  $Tk \times 1$  covariance  $\sigma_{x\alpha}$  between the random intercept and all true regressors  $\mathbf{x}_i$  can be estimated. Thus, various forms of time-dependent errors can be accommodated. One may postulate heteroscedastic measurement errors, i.e.,

$$V[\mathbf{u}_i^{(j)}] = V \left[ \begin{pmatrix} u_{i1}^{(j)} \\ u_{i2}^{(j)} \\ \vdots \\ u_{iT}^{(j)} \end{pmatrix} \right] = \begin{bmatrix} \sigma_{j1} & 0 & \dots & 0 \\ 0 & \sigma_{j2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{jT} \end{bmatrix}, \quad j = 1, 2, \dots, k.$$

Measurement errors may also be assumed to follow some autoregressive (e.g., AR(1) or AR( $T-2$ )) or some moving average processes. The same is true for equation errors,  $e_{it}$ . As in Theorem 2, if  $\beta_j = 0$  for some  $j$ , only the corresponding  $\Sigma_{uu_j}$  is not identified.

Note that Theorems 1, 2 and 3 list sufficient conditions for identification. By satisfying these conditions, one can ensure that the relevant model parameters are identified. However, some model parameters may still be identified even if the conditions given in these theorems are violated.

## Estimation

We consider a method for fitting model (3) that is based on the first two sample moments and that does not specify a particular distributional form. For the observations  $\mathbf{Z}_i$  in (5), let

$$\bar{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i,$$

and

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})'$$

If  $\mathbf{Z}_i$  is normally distributed, then  $\bar{\mathbf{Z}}$  and  $\mathbf{S}$  form a set of sufficient statistics, and the maximum likelihood estimator of  $\theta$  in model (3) can be obtained by minimizing

$$F(\theta) = n\{[\bar{\mathbf{Z}} - \gamma(\theta)]'\Omega^{-1}(\theta)[\bar{\mathbf{Z}} - \gamma(\theta)] + \ln |\Omega(\theta)| + \text{tr}[\mathbf{S}\Omega^{-1}(\theta)] - \ln |\mathbf{S}| - T(k+1)\}. \quad (16)$$

Let  $\hat{\theta}$  denote the estimator obtained in this way. When  $\mathbf{Z}_i$  is non-normal,  $\hat{\theta}$  is still reasonable from the point of view of considering the function  $F(\theta)$  in (16) as a discrepancy between the sample moments  $(\bar{\mathbf{Z}}, \mathbf{S})$  and the population moments  $(\gamma(\theta), \Omega(\theta))$ . No assumption other than the existence of the first two moments of the observed variables is needed. In this case,  $\hat{\theta}$  is a method of moments (or moment structure analysis) estimator for model (3) without distributional assumptions. Another justification for using  $\hat{\theta}$  for non-normal data is given in the next section.

There are various ways of testing the overall fit of the model. One straightforward test can be derived directly from the value of the discrepancy function in (16). If  $\mathbf{Z}_i$  is normally distributed, the likelihood ratio test for testing the form of  $\gamma(\theta)$  and  $\Omega(\theta)$  as specified in model (3) against unrestricted first two moments is to reject the null hypothesis when  $F(\hat{\theta})$  exceeds an upper percentile of a chi-square distribution. The degrees of freedom  $q$  for the chi-square distribution is given by subtracting the number of parameters in  $\theta$  from the number of elements in the unrestricted  $\gamma$  and  $\Omega$ .

$$T(k+1) + \frac{1}{2}T(k+1)[T(k+1) + 1].$$

In the next section, the use of this goodness-of-fit test for model (3) is considered when  $\mathbf{Z}_i$  is non-normal.

The estimator  $\hat{\theta}$  can be computed using the standard statistical packages for moment structure analysis or structural equation modeling, e.g., LISREL (Jöreskog and Sörbom, 1983) and PROC CALIS (SAS Institute, 1991). Even packages that perform covariance structure analysis, i.e., fit  $\Omega(\theta)$  to  $\mathbf{S}$  by minimizing

$$n\{\ln |\Omega(\theta)| + \text{tr}[\mathbf{S}\Omega^{-1}(\theta)] - \ln |\mathbf{S}| - T(k+1)\}, \quad (17)$$



can be used to obtain  $\hat{\theta}$  and  $F(\hat{\theta})$  by substituting the augmented moment matrices

$$S^* = \begin{bmatrix} S + \bar{Z}\bar{Z}' & \bar{Z}' \\ \bar{Z} & 1 \end{bmatrix} \quad (18)$$

and

$$\Omega^*(\theta) = \begin{bmatrix} \Omega(\theta) + \gamma(\theta)\gamma(\theta)' & \gamma(\theta)' \\ \gamma(\theta) & 1 \end{bmatrix} \quad (19)$$

for  $S$  and  $\Omega(\theta)$  in (17) to yield an equivalent form of (16) (Satorra, 1992).<sup>1</sup> Such packages also compute asymptotic standard errors for all parameter estimates under the assumption of normal  $Z_i$ . The validity and usefulness of such standard errors in the case of non-normal  $Z_i$  are addressed in the next section.

### Asymptotic Properties

In applications of the random effect analysis to panel data, the normal or even random sample assumption is often untenable. Observations may be discrete, bounded or from skewed distributions, and therefore, clearly non-normal. However, the estimator  $\hat{\theta}$ , its standard error under normality, and the test of model fit  $F(\hat{\theta})$  turn out to be useful for most non-normal data.

Estimation when the normality assumption does not in fact hold is often done in practice and is referred to in the literature as pseudo maximum likelihood (PML). Despite the violation of the normality assumption underlying the use of (16), the resulting PML estimator of  $\theta$  is consistent. However, inferences based on PML statistics may not be valid. Browne (1984) suggested using asymptotic distribution-free (ADF) methods, but ADF methods tend to be computationally intensive and statistically unstable since they involve fourth-order moments. Studies have also found that very large samples are required to get the ADF-based chi-square goodness-of-fit test to perform adequately (see, e.g., Hu *et al.*, 1992 and West *et al.*, 1995). In a different line of research, researchers have demonstrated that statistics generated under the normality assumption yield valid inferences even when the normality assumption cannot be made (see Browne, 1987; Anderson and Amemiya, 1988; Browne and Shapiro, 1988; Amemiya and Anderson, 1990; Satorra, 1992; and Papadopoulos and Amemiya, 1996). The general results in these papers are applicable for our model. Thus, we state, without proof, the asymptotic properties of our estimator  $\hat{\theta}$  and the goodness-of-fit test statistic  $F(\hat{\theta})$ .

The primary interest here is in the asymptotic covariance matrix of  $\hat{\theta}$  and the asymptotic distribution of  $F(\hat{\theta})$ . Thus, throughout this section, we assume that identification conditions are met so that a model

<sup>1</sup> The error degrees of freedom reported by the package needs to be decreased by one to obtain the correct value. This adjustment must be made since  $\Omega^*(\theta)$  in (19) has one non-informative element that is a constant equal to one.

being discussed is identified. Then, the consistency and asymptotic normality of  $\hat{\theta}$  readily follows. It turns out that all of our results hold for essentially any type of random or fixed true values  $\mathbf{x}_{it}$ . For this, we assume:

**Assumption A** *The independent and identically distributed  $\alpha_i$  and the random or fixed true values  $\mathbf{x}_i$  satisfy, as  $n \rightarrow \infty$ ,*

$$\begin{pmatrix} \bar{\alpha} \\ \text{vec } \bar{\mathbf{x}} \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \alpha_i \\ \text{vec } \mathbf{x}_i \end{pmatrix} \rightarrow \begin{pmatrix} \mu_{\alpha} \\ \mu_x \end{pmatrix}, \text{ a.s.},$$

$$\frac{1}{n} \sum_{i=1}^n \begin{bmatrix} \alpha_i - \bar{\alpha} \\ \text{vec } (\mathbf{x}_i - \bar{\mathbf{x}}) \end{bmatrix} \begin{bmatrix} \alpha_i - \bar{\alpha}, & [\text{vec } (\mathbf{x}_i - \bar{\mathbf{x}})]' \end{bmatrix} \rightarrow \Omega, \text{ a.s.},$$

where

$$\Omega = \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma'_{x\alpha} \\ \sigma_{x\alpha} & \Sigma_{xx} \end{bmatrix}$$

is nonsingular.

In the following three results (Theorems 4-6) concerning the asymptotic inference procedures, the required conditions involve the class of models being fitted. Therefore, we assume throughout that the true model is contained in the class of models being fitted. For instance, the condition that a diagonal covariance matrix is fitted implicitly assumes that the true covariance is a diagonal matrix (although all diagonal elements may be equal).

The first result assumes the relatively strong assumption of normal  $\mathbf{e}_i$  and  $\mathbf{u}_i^{(j)}$  but allows any unrestricted or time structure for  $\mathbf{e}_i$  and  $\mathbf{u}_i^{(j)}$ , provided the model, with or without  $\sigma_{x\alpha} = 0$ , is identified.

**Theorem 4** *Let Assumption A hold. Assume that  $\mathbf{e}_i$  and  $\mathbf{u}_i^{(j)}$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, k$ , are normally distributed and that the model given by (3) is identified. Then, the asymptotic covariance matrix of  $(\hat{\mu}_{\alpha}, \hat{\beta}')'$  is the same as that obtained under the normality of  $\mathbf{Z}_i$ , and  $F(\hat{\theta})$  converges in distribution to  $\chi_q^2$ .*

Thus, for any distribution of  $\alpha_i$ , for any type of random or fixed  $\mathbf{x}_i$ , and for any identified model, asymptotic inferences for  $\mu_{\alpha}$  and  $\beta$  and for model fit made under the incorrect assumption of normally distributed observations are valid, provided the errors are normal. The normal error assumption can also be weakened keeping the same conclusion. For the case where  $\sigma_{x\alpha}$  is known to be zero and is not estimated, we have the next result.

**Theorem 5** *Suppose  $\sigma_{x\alpha}$  is known to be zero, and that Assumption A holds. Assume either*

(a)  $e_i$  is normally distributed, and any restricted or unrestricted  $\Sigma_{ee}$  is fitted (provided model (3) is identified),

or

(b)  $e_i$  has any distribution, and a restricted  $\Sigma_{ee}$ , specifically

$$\Sigma_{ee} = \begin{bmatrix} \sigma_{ee,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{ee,T} \end{bmatrix}, \quad (20)$$

is fitted.

Assume also that the  $u_i^{(j)}$ 's have any distribution, and assume that for each  $j = 1, 2, \dots, k$ , either

(i) an unrestricted  $\Sigma_{uu_j}$  is fitted,

or

(ii) the heteroscedastic structure

$$\Sigma_{uu_j} = \begin{bmatrix} \sigma_{j1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{jT} \end{bmatrix} \quad (21)$$

is fitted.

Then, the asymptotic covariance matrix of  $(\hat{\mu}_\alpha, \hat{\beta}')'$  is the same as that under the normality of  $Z_i$ , and  $F(\hat{\theta})$  converges in distribution to  $\chi_q^2$ .

For the case when  $\sigma_{x\alpha}$  is to be estimated, the model may not be identified if the whole  $\Sigma_{uu}$  has to be estimated. Thus, we have the following result.

**Theorem 6** Suppose that  $\sigma_{x\alpha}$  is to be estimated, and that Assumption A holds. Assume either

(a)  $e_i$  is normally distributed, and any restricted or unrestricted  $\Sigma_{ee}$  is fitted (provided model (3) is identified),

or

(b)  $e_i$  has any distribution, and a restricted  $\Sigma_{ee}$ , specifically

$$\Sigma_{ee} = \begin{bmatrix} \sigma_{ee,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{ee,T} \end{bmatrix},$$

is fitted.

Assume also that the  $u_i^{(j)}$ 's have any distribution, and assume that for each  $j = 1, 2, \dots, k$ , the heteroscedastic structure

$$\Sigma_{uu_j} = \begin{bmatrix} \sigma_{j1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{jT} \end{bmatrix}$$

is fitted. Then, the asymptotic covariance matrix of  $(\hat{\mu}_\alpha, \hat{\beta})'$  is the same as that under normality of  $Z_i$ , and  $F(\hat{\theta})$  converges in distribution to  $\chi_q^2$ .

Hence, the asymptotic inferences for  $(\mu_\alpha, \beta)'$  and for model fit can be carried out correctly using the normality-based statistical packages for a very broad class of non-normal data. If a diagonal covariance matrix is being fitted to each error term, then any distributional form is acceptable for every variable involved in the model. The diagonal heteroscedastic form is often reasonable for measurement errors  $u_i^{(j)}$ . If  $\sigma_{x\alpha}$  is treated as zero, the unrestricted  $\Sigma_{uu}$  can be fitted without losing the validity of the normality-based inferences. The normality of  $e_{it}$ , the error in the equation, is considered a reasonable assumption in most cases, but this can also be dropped according to Theorems 5 and 6. Note that there is no restriction on the distributional form of the random intercept  $\alpha_i$  and the  $T \times k$  true explanatory variable  $x_i$  in Theorems 5 and 6, as long as Assumption A holds. In fact,  $x_i$  does not need to form a random sample and can be considered fixed constants. In some applications, a reasonable model for  $e_i$  and  $u_i^{(j)}$  is independent and identical distributions over time with a covariance matrix proportional to the  $T \times T$  identity matrix, but the normality of the error term is suspect. In such a situation, the heteroscedastic covariance matrix with  $T$  diagonal elements to be estimated can be fitted to obtain valid normality-based asymptotic inferences. The loss of efficiency in fitting a larger number of parameters than needed is small for the estimation of  $\mu_\alpha$  and  $\beta$  because the contribution of estimating error structure is small in the overall variability of  $\hat{\mu}_\alpha$  and  $\hat{\beta}$ .

## Simulation Study

This study consists of two parts. For both parts, model (3) with  $k = 1$ ,  $T = 3$  and  $n = 200$  is used. In the first part, the proposed pseudo maximum normal likelihood (ML) estimator and the Griliches and Hausman (1986) instrumental variable estimator (IV/GMM) of  $\beta$  are compared when observations are normally distributed. The second part assesses the finite sample properties of the inference procedures based on the pseudo maximum normal likelihood method when observations are not normally distributed. Throughout, the ML computation was carried out using the SAS CALIS procedure using the Cholesky decomposition parameterization for any unrestricted covariance matrix.

### IV/GMM Estimation versus ML Estimation

Data were generated according to model (3) in which all random variables, including  $\alpha_i$ ,  $x_{it}$ ,  $u_{it}$  and  $e_{it}$  were normally distributed. The true values of the parameters were set to

$$\beta = 1, \quad \mu_\alpha = 4, \quad \sigma_{\alpha\alpha} = 1, \quad \sigma_{x\alpha} = 0, \quad \text{and} \quad \Sigma_{ee} = 2\mathbf{I}_3. \quad (22)$$

For the true values of  $\mu_x$ ,  $\Sigma_{xx}$  and  $\Sigma_{uu}$ , we considered three sets as described later. For the ML estimation, model (3), with  $\sigma_{x\alpha}$  estimated,  $\Sigma_{ee} = \sigma_{ee}\mathbf{I}_3$ , and

$$\Sigma_{uu} = \begin{bmatrix} \sigma_{11} & \sigma_{21} & 0 \\ \sigma_{21} & \sigma_{22} & \sigma_{32} \\ 0 & \sigma_{32} & \sigma_{33} \end{bmatrix}, \quad (23)$$

was fitted to each sample via maximum normal likelihood. Estimates of  $\beta$  were also computed using the Griliches and Hausman (1986) IV/GMM method. The IV/GMM estimation depends on the choice of instruments, which in turn, heavily relies on the assumed knowledge concerning the behavior of  $x_{it}$  and  $u_{it}$ . Throughout, we used the instruments as given in Table 1. These instruments would all be valid under the assumed knowledge that  $u_{it}$ 's are independent and identically distributed over time and individuals, that the mean of the true  $x_{it}$  differs over the three time points, that the three variances of  $x_{it}$ ,  $t = 1, 2, 3$ , are different, and that the three covariances among  $x_{it}$ ,  $t = 1, 2, 3$  are unequal (unless a certain special structure holds). First, the two-stage least squares procedure is applied to each of the three equations with the two instruments to obtain an initial estimate of  $\beta$ . Then, the three equations are combined by the generalized method of moments method to obtain the overall IV/GMM estimate of  $\beta$ ,  $\tilde{\beta}_{\text{GMM}}$ , as described in Griliches and Hausman (1986). In addition to the  $\tilde{\beta}_{\text{GMM}}$  defined by Griliches and Hausman (1986), a modified version of  $\tilde{\beta}$ , denoted by  $\tilde{\beta}_{\text{GMM}}^*$ , was computed. The modification involved using a more efficient weight matrix arising from the assumption that both  $e_{it}$  and  $u_{it}$  are independent and identically distributed over  $i$  and  $t$ .

Table 1 Instruments used to compute initial estimates of  $\beta$

Equation to be estimated	Instruments
$Y_{i2} - Y_{i1} = \beta(x_{i2} - x_{i1}) + (\epsilon_{i2} - \epsilon_{i1})$	$X_{i1} + X_{i2}, X_{i3}$
$Y_{i3} - Y_{i2} = \beta(x_{i3} - x_{i2}) + (\epsilon_{i3} - \epsilon_{i2})$	$2(X_{i2} + X_{i3}), X_{i1}$
$Y_{i3} - Y_{i1} = \beta(x_{i3} - x_{i1}) + (\epsilon_{i3} - \epsilon_{i1})$	$X_{i1} + X_{i3}, X_{i2}$

The first set of 1000 samples was generated according to model (3) with true parameter values (22) and, in addition,

$$\mu_x = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \quad \Sigma_{xx} = \begin{bmatrix} 4.0 & 2.0 & 0.0 \\ 2.0 & 4.0 & 0.0 \\ 0.0 & 0.0 & 4.0 \end{bmatrix}, \quad \text{and} \quad \Sigma_{uu} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

The ML and IV/GMM estimates of  $\beta$  were computed as described above. Box plots of the resulting 1000 estimates are given in Figure 1, where the dotted horizontal line marks the true value of the parameter. The bias, variance and mean squared error for the three estimators are given in Table 2. For these true values, the assumed conditions for the instruments in Table 1 used in the IV/GMM estimation are not satisfied. In fact, for each of the three equations, both instruments used are invalid. As a result, the IV/GMM and modified IV/GMM estimators of  $\beta$  have large bias and mean squared error (Table 2) and almost always take values less than the true value (Figure 1). This illustrates the fact that if the specific knowledge assumed for the instrument choice is incorrect, then the performance of the IV/GMM estimator can be very poor and unsatisfactory. On the other hand, the ML estimator gives a reasonable sampling distribution even with estimation of  $\sigma_{x\alpha}$  and fitting many parameters in  $\Sigma_{uu}$  of the form (23).

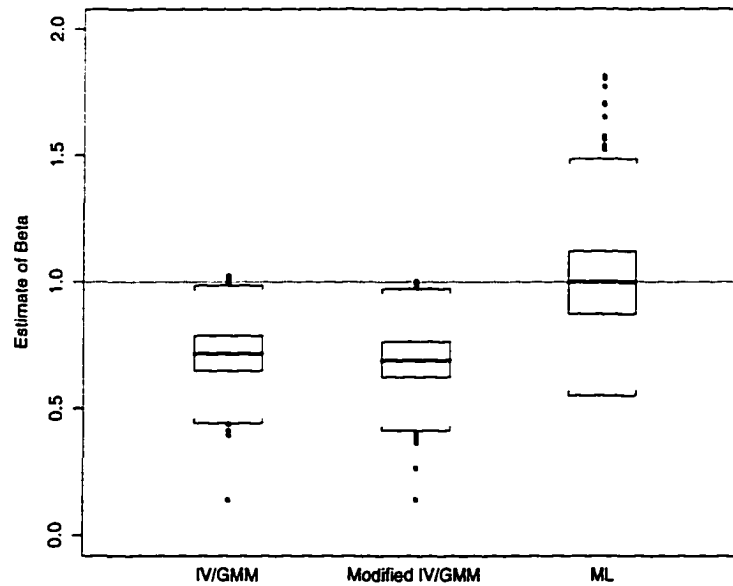


Figure 1 Box plots of estimates of  $\beta$  for case 1

Table 2 Monte Carlo moments for case 1

Estimator	Bias	Variance	MSE
$\tilde{\beta}_{\text{GMM}}$	-0.28382	0.01097	0.09152
$\tilde{\beta}_{\text{GMM}}^*$	-0.30778	0.01243	0.10716
$\hat{\beta}_{\text{ML}}$	0.00450	0.03557	0.03559

In the second case, the true values of  $\mu_x$ ,  $\Sigma_{xx}$ , and  $\Sigma_{uu}$  were set to be

$$\mu_x = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \quad \Sigma_{xx} = \begin{bmatrix} 4.0 & 0.0 & 0.0 \\ 0.0 & 4.8 & 0.0 \\ 0.0 & 0.0 & 5.6 \end{bmatrix}, \quad \text{and} \quad \Sigma_{uu} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \quad (24)$$

Then, 1000 samples were generated with the true values (22) and (24). The results on the three estimators are summarized in Figure 2 and Table 3. For this case, the assumed knowledge for the instruments in Table 1 used for the IV/GMM estimation is partially incorrect. That is, for each of the three equations, only the first of the two instruments is informative while both instruments are valid in the sense of independence with the error term. As exhibited in Figure 2 and Table 3, the IV/GMM and modified IV/GMM estimators possess sizable biases and tend to take many outlying values away from the true value of  $\beta$ . The ML estimator has a much tighter sampling distribution around the true value. Thus, the IV/GMM estimator with a less than ideal set of instruments can produce a heavy-tailed sampling distribution and performs much worse than the ML estimator that estimates a larger number of parameters under less restrictive assumptions.

Table 3 Monte Carlo moments for case 2

Estimator	Bias	Variance	MSE
$\tilde{\beta}_{\text{GMM}}$	-0.21359	0.10406	0.14968
$\tilde{\beta}_{\text{GMM}}^*$	-0.21124	0.12188	0.16650
$\hat{\beta}_{\text{ML}}$	0.12163	0.08347	0.09827

The third set of 1000 samples was generated using the true values in (22) and

$$\mu_x = \begin{pmatrix} 2 \\ 5 \\ 10 \end{pmatrix}, \quad \Sigma_{xx} = \begin{bmatrix} 4.0 & 2.0 & 0.8 \\ 2.0 & 4.8 & 2.8 \\ 0.8 & 2.8 & 5.6 \end{bmatrix}, \quad \Sigma_{uu} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \quad (25)$$

This is the case where all the knowledge assumed for the IV/GMM estimation is in fact correct and where the instruments in Table 1 provide an ideal set. The simulation results are given in Table 4 and

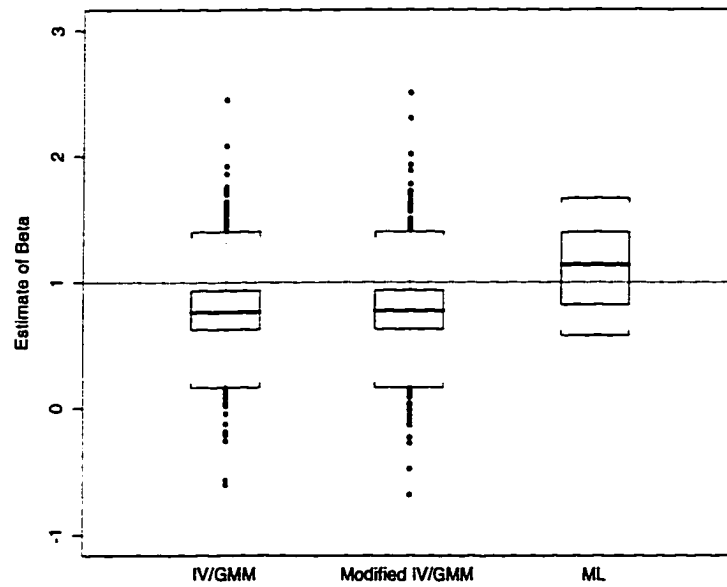


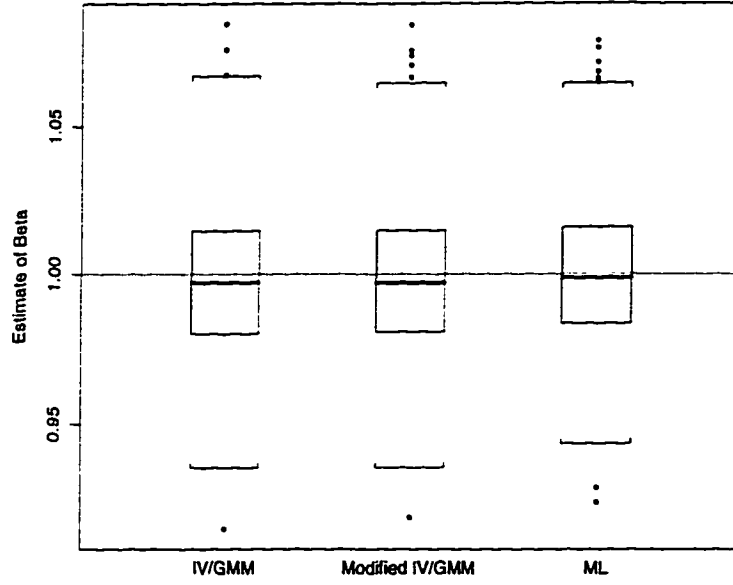
Figure 2 Box plots of estimates of  $\beta$  for case 2  
(Two observations,  $\tilde{\beta}_{\text{GMM}} = 5.46$  and  $\tilde{\beta}_{\text{GMM}}^* = 6.65$  fall outside the range of values presented in this figure.)

Figure 3. Table 4 also lists the bias and mean squared error of another estimator of  $\beta$ ,  $\tilde{\beta}$ , obtained by the maximum likelihood method ignoring the existence of measurement error in the explanatory variable. This estimator has a small variance but a very large bias, and thus, it is an unacceptable estimator of  $\beta$ . Estimation of  $\beta$  ignoring measurement errors can lead to a serious error, even when the measurement error variances are only  $\frac{1}{4}$  to  $\frac{1}{3}$  of the total variability of observed explanatory variables. In this case, with very informative instruments, the IV/GMM and the modified IV/GMM estimators perform nearly as efficiently as the ML estimator. However, even in such a situation, the ML method that estimates more parameters under more general conditions produces an estimator of  $\beta$  with smaller bias and variance than the IV/GMM estimators.

Table 4 Monte Carlo moments for case 3

Estimator	Bias	Variance	MSE
$\tilde{\beta}$	-0.10593	0.00036	0.01160
$\tilde{\beta}_{\text{GMM}}$	-0.00201	0.00062	0.00062
$\tilde{\beta}_{\text{GMM}}^*$	-0.00189	0.00060	0.00061
$\tilde{\beta}_{\text{ML}}$	0.00028	0.00057	0.00057



Figure 3 Box plots of estimates of  $\beta$  for case 3

### Finite Sample Properties of the ML Estimator

Simulations were also conducted to assess the finite sample properties of the maximum normal likelihood estimator when observations are not normal. The first set of results deals with situations in which the equation and measurement errors are normal but the true  $x$ 's and the random effect  $\alpha_i$  are non-normal. For these cases, by Theorem 4, the standard error estimates and chi-square tests of model fit obtained under the normality of observations are still valid, although observations are not normal. Data were generated according to model (3) with  $k = 1$ ,  $T = 3$ ,  $n = 200$ , and true parameter values (22) and (25). The equation errors  $e_{it}$  and the measurement errors  $u_{it}$  were normally distributed. Four different sets of 1000 samples were generated with different distributions of  $x_{it}$  and  $\alpha_i$  listed in Tables 5 to 6, where N indicates a normal distribution and  $\chi^2$  indicates a shifted and scaled  $\chi^2$  distribution with two degrees of freedom. The first set with normal  $x_i$  and normal  $\alpha_i$  actually gives normal observations. Two versions of model (3) were fitted to each sample via maximum normal likelihood. In one version,  $\sigma_{x\alpha} = 0$  was not estimated, an unrestricted  $3 \times 3$  matrix  $\Sigma_{uu}$  was fitted, and  $\Sigma_{ee}$  was fitted as  $\sigma_{ee}\mathbf{I}_3$ . The second version involved estimating  $\sigma_{x\alpha}$ , fitting  $\Sigma_{ee}$  as  $\sigma_{ee}\mathbf{I}_3$ , and fitting  $\Sigma_{uu}$  as  $\sigma_{uu}\mathbf{I}_3$ . The goodness-of-fit statistic  $F(\hat{\theta})$  as well as 95% confidence intervals for  $\mu_\alpha$  and for  $\beta$  under pseudo normality were obtained for each sample and for each method of fitting model (3). Table 5 reports the simulated coverage probabilities while Table 6 presents the simulated probabilities for  $F(\hat{\theta})$  to exceed

the 0.10, 0.05 and 0.01 upper quantiles of the appropriate  $\chi^2$  distribution. These tables indicate that Theorem 4 holds very well in finite samples and for various non-normal distributions for the true  $x$ 's and the random effect  $\alpha_i$ . The coverage probabilities in Table 5 are all similar to the nominal value of 0.95, and the upper percentiles of the fit statistic are well approximated by the  $\chi^2$  percentiles. Thus, valid inferences about  $\mu_\alpha$  and  $\beta$  and about model fit can be made using standard errors and chi-square test statistics computed under the pseudo normality of observations for non-normal distributions of the random effect  $\alpha_i$  and the true explanatory variables  $x_i$  when the equation and measurement errors are normally distributed.

Table 5 Simulated coverage probabilities of 95% confidence intervals (normal errors)

Distribution		$\sigma_{x\alpha} = 0$		$\sigma_{x\alpha}$ estimated	
$x_i$	$\alpha_i$	$\mu_\alpha$	$\beta$	$\mu_\alpha$	$\beta$
N	N	0.953	0.952	0.953	0.955
$\chi^2$	N	0.954	0.951	0.950	0.953
N	$\chi^2$	0.963	0.952	0.957	0.954
$\chi^2$	$\chi^2$	0.943	0.948	0.948	0.947

Table 6 Simulated frequency of rejection of the  $\chi^2$  goodness-of-fit test (normal errors)

Distribution		$\sigma_{x\alpha} = 0$			$\sigma_{x\alpha}$ estimated		
$x_i$	$\alpha_i$	10%	5%	1%	10%	5%	1%
N	N	0.098	0.053	0.008	0.111	0.050	0.014
$\chi^2$	N	0.089	0.053	0.010	0.086	0.040	0.006
N	$\chi^2$	0.106	0.052	0.014	0.118	0.069	0.017
$\chi^2$	$\chi^2$	0.097	0.049	0.011	0.103	0.052	0.014

When the normality assumption cannot be made for both the equation errors and the measurement errors, the restrictions on the fitted error covariance matrices in Theorems 5 and 6 need to be satisfied for the normality-based inferences to remain valid. For this, data were generated according to model (3) with true parameter values (22) and (25). For  $x_i$ , a normal sample with  $n = 200$  was obtained, and the same set was used for all 1000 simulation samples. Thus,  $x_i$ 's were fixed over repeated samples. Since all other variables are independent of  $x_i$ , the fixed  $x_i$  experiment corresponds to studying the sampling distribution conditional on  $x_i$ 's. Thus, the coverage probability and percentile results here would apply to any  $x_i$  that are independent of other random components of the model. Tables 7 and

8 indicate normal (N) and  $\chi^2$  with two degrees of freedom ( $\chi^2$ ) distributions used for  $\alpha_i$ ,  $u_{it}$ , and  $e_{it}$ , where  $(u_{it}, e_{it})$  are all independent and identically distributed over  $t$  under (22) and (25). Two models were fitted to each sample with and without fitting  $\sigma_{x\alpha}$ . For the model with  $\sigma_{x\alpha}$  not estimated,  $\Sigma_{ee}$  in (20) was fitted while  $\Sigma_{uu}$  was fitted as an unrestricted  $3 \times 3$  covariance matrix. For the model with  $\sigma_{x\alpha}$  estimated,  $\Sigma_{ee}$  in (20) and the diagonal  $\Sigma_{uu}$  in (21) were fitted. Both models satisfy the conditions in Theorems 4 and 5. The results are summarized in Tables 7 and 8. The coverage probabilities in Table 7 are uniformly close to the nominal 95%, and the  $\chi^2$  approximation to the model fit statistic is very good in upper percentiles as indicated in Table 8. Thus, with proper application of the ML fitting based on Theorems 4 - 6, accurate asymptotic inferences can be made regarding  $(\mu_\alpha, \beta)'$  and model fit for a broad class of non-normal data.

Table 7 Simulated coverage probabilities of 95% confidence intervals  
( $\mathbf{x}_i$  fixed over repeated sampling)

Distribution			$\sigma_{x\alpha} = 0$		$\sigma_{x\alpha}$ estimated	
$\alpha_i$	$u_{it}$	$e_{it}$	$\mu_\alpha$	$\beta$	$\mu_\alpha$	$\beta$
N	N	N	0.948	0.949	0.948	0.945
$\chi^2$	N	N	0.953	0.951	0.951	0.951
N	$\chi^2$	N	0.952	0.957	0.950	0.959
N	N	$\chi^2$	0.956	0.950	0.961	0.956
$\chi^2$	$\chi^2$	N	0.945	0.945	0.951	0.950
$\chi^2$	N	$\chi^2$	0.947	0.942	0.951	0.949
N	$\chi^2$	$\chi^2$	0.948	0.955	0.953	0.959
$\chi^2$	$\chi^2$	$\chi^2$	0.954	0.937	0.951	0.942

Table 8 Simulated frequency of rejection of the  $\chi^2$  goodness-of-fit test  
( $\mathbf{x}_i$  fixed over repeated sampling)

Distribution			$\sigma_{x\alpha} = 0$			$\sigma_{x\alpha}$ estimated		
$\alpha_i$	$u_{it}$	$e_{it}$	10%	5%	1%	10%	5%	1%
N	N	N	0.112	0.053	0.008	0.111	0.055	0.006
$\chi^2$	N	N	0.112	0.043	0.011	0.101	0.052	0.009
N	$\chi^2$	N	0.093	0.050	0.011	0.091	0.048	0.007
N	N	$\chi^2$	0.087	0.050	0.011	0.096	0.055	0.009
$\chi^2$	$\chi^2$	N	0.107	0.054	0.012	0.099	0.052	0.010
$\chi^2$	N	$\chi^2$	0.099	0.057	0.003	0.101	0.047	0.008
N	$\chi^2$	$\chi^2$	0.105	0.043	0.010	0.099	0.051	0.008
$\chi^2$	$\chi^2$	$\chi^2$	0.114	0.049	0.009	0.105	0.053	0.011

## Summary

Random effect analysis with measurement error has not been widely studied, yet it is important considering the increasing use of panel data in recent years. The existing estimation procedure for such a model involves instrumental variables and is difficult to use in practice. This paper proposed a unified approach to the analysis that provides a systematic way of checking the identification of the model as well as a method for fitting the model. The model fitting method based on pseudo maximum normal likelihood gives inference procedures useful for a wide range of non-normal data. The simulation study supported the usefulness of the approach in practice.

## References

- Aasness, K., E. Biørn and T. Skjerpen, 1993, Engel functions, panel data, and latent variables, *Econometrica* 61, 1395-1422.
- Altonji, J. G. and A. Siow, 1987, Testing the response of consumption to income changes with (noisy) panel data, *Quarterly Journal of Economics* 102, 293-328.
- Amemiya, Y. and T. W. Anderson, 1990, Asymptotic chi-square tests for a large class of factor analysis models, *Annals of Statistics* 18, 1453-1463.
- Anderson, T. W. and Y. Amemiya, 1988, The asymptotic normal distribution of estimators in factor analysis under general conditions, *Annals of Statistics* 16, 757-771.
- Biørn, E., 1992, Panel data with measurement errors, in: L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of Theory and Applications* (Kluwer Academic Publishers, Dordrecht).
- Bollen, K. A., 1989, *Structural Equations with Latent Variables* (John Wiley, New York).
- Bollen, K. A. and J. S. Long, 1993, *Testing Structural Equation Models* (Sage Publications, Thousand Oaks).
- Bound, J., C. Brown, G. J. Duncan and W. L. Rogers, 1990, Measurement error in cross-sectional and longitudinal labor market surveys: validation study evidence, in: J. Hartog, G. Ridder and T. Theeuwes, eds., *Panel Data and Labor Market Studies* (Elsevier Science, New York).
- Bowers, N. and F. W. Horvath, 1984, Keeping time: an analysis of errors in the measurement of unemployment duration, *Journal of Business and Economic Statistics* 2, 140-149.
- Browne, M. W., 1984, Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology* 37, 62-83.
- Browne, M. W., 1987, Robustness of statistical inference in factor analysis and related models, *Biometrika* 74, 375-384.

- Browne, M. W. and A. Shapiro, 1988, Robustness of normal theory methods in the analysis of linear latent variate models, *British Journal of Mathematical and Statistical Psychology* 41, 193-208.
- Duncan, G. J. and D. H. Hill, 1985, An investigation of the extent and consequences of measurement error in labor economic survey data, *Journal of Labor Economics* 3, 508-532.
- Fuller, W. A., 1987, *Measurement Error Models* (John Wiley, New York).
- Griliches, Z. and J. A. Hausman, 1986, Errors in variables in panel data, *Journal of Econometrics* 31, 93-118.
- Hansen, L., 1982, Large sample properties of generalized method of moments, *Econometrica* 50, 1029-54.
- Hoyle, R., 1995, *Structural Equation Models: Concepts, Issues and Applications* (Sage Publications, Thousand Oaks).
- Hsiao, C. and G. Taylor, 1991, Some remarks on measurement error and the identification of panel data models, *Statistica Neerlandica* 45, 187-194.
- Jöreskog, K. G. and D. Sorbom, 1983, *LISREL V and LISREL VI : analysis of linear structural relationships by maximum likelihood and least squares methods* (International Educational Services, Chicago).
- Mundlak, Y., 1978, On the pooling of cross-section and time-series data, *Econometrica* 46, 69-85.
- Papadopoulos, S. and Y. Amemiya, 1996, On linear latent variable analysis of multiple populations, Preprint Series No. 96-31, Statistical Laboratory, Iowa State University.
- SAS Institute, Inc., 1991, *The CALIS procedure: analysis of covariance structures* (Author, Cary, NC).
- Satorra, A., 1992, Asymptotic robust inferences in the analysis of mean and covariance structures, in: P. V. Marsden, ed., *Sociological Methodology*.
- Solon, G. S., 1989, The value of panel data in economic research, in D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh, eds., *Panel Surveys* (John Wiley, New York).
- Wansbeek, R. J. and R. H. Koning, 1989, Measurement error and panel data, *Statistica Neerlandica* 45, 85-92.
- West, S. G., J. F. Finch and P. J. Curran, Structural equation models with nonnormal data, in: R. Hoyle, ed., *Structural Equation Models: Concepts, Issues and Applications* (Sage Publications, Thousand Oaks).

## RANDOM COEFFICIENT REGRESSION WITH ERRORS IN VARIABLES

A paper to be submitted to *Econometric Theory*

Elizabeth Martha Paterno and Yasuo Amemiya

### Abstract

A random coefficient model that accounts for measurement error in the explanatory variables is studied. Two procedures are proposed for model fitting and estimation. The generalized least squares method is developed for the first two sample moments with a distribution-free estimate of the weight. Since this method tends to yield very variable estimates in small samples, an alternative method, the pseudo maximum normal likelihood procedure is also developed. The latter, obtained by maximizing a hypothetical normal likelihood for the first two sample moments, produces relative stable estimates in most samples. Asymptotic properties of the two procedures are derived and are used to obtain valid standard errors of the estimators. Numerical results showing the finite-sample properties of these estimators are also reported.

### Introduction

A panel data set consists of measurements taken over time from several individuals. Typically, a dependent response variable  $Y_{it}$  and a covariate vector  $\mathbf{x}_{it}$  are considered for  $i = 1, 2, \dots, n$ ,  $t = 1, 2, \dots, T$ , where  $n$  is the number of individuals,  $T$  is the number of time points common to all individuals and the vector  $\mathbf{x}_{it}$  includes time- and individual-dependent covariates. In economics and other social sciences, a rectangular panel data set with no missing values is usually available and typically has a short time series on a large number of individuals. Often, researchers seek to describe the relationship between  $\mathbf{x}_{it}$  and  $Y_{it}$  by fitting linear models of the type

$$Y_{it} = \alpha_0 + \beta_0' \mathbf{x}_{it} + e_{it}^0.$$

where the  $e_{it}^0$ 's represent random errors with mean zero. Since observations taken from the same individual are potentially correlated, most models fitted to panel data attempt to account for the within-individual homogeneity and between-individual heterogeneity in various ways. One such model is the random coefficient model that assumes that individual heterogeneity is manifested in intercepts and slopes that are stochastic and different over individuals. This is considered an appropriate assumption when individuals from whom measurements are taken are randomly sampled from some population. The random coefficient model contains, as special cases, the fixed and mixed coefficient models with some or all coefficients common for all individuals. The model where only the intercept is random over individuals is the standard random effect model. Consider the general random coefficient model

$$\begin{aligned} Y_{it} &= \beta_i' \begin{pmatrix} 1 \\ \mathbf{x}_{it} \end{pmatrix} + e_{it}, \\ \beta_i &\sim (\beta, \Phi), \end{aligned} \quad (1)$$

where  $\beta_i$  is independent of  $e_{it}$ ,  $\beta$  is the average coefficient, and  $\Phi$  represents the individual variability in all coefficients. This model provides a very flexible way of accounting for individual variability in panel data regression. As opposed to fitting different parameters  $\beta_i$  for all individuals, the random coefficient specification reduces the number of parameters to be estimated while still accounting for individual heterogeneity by allowing coefficients to differ across individuals.

In many applications, however, measurement error is often part of what is observed of  $\mathbf{x}_{it}$ . That is,  $\mathbf{X}_{it}$  is observed, where

$$\mathbf{X}_{it} = \mathbf{x}_{it} + \mathbf{u}_{it}. \quad (2)$$

Here, the measurement error  $\mathbf{u}_{it}$  is assumed to have mean zero and to be independent of  $\mathbf{x}_{it}$  and  $e_{it}$ . Measurement errors for different components of  $\mathbf{x}_{it}$  can often be treated as independent, but errors may be correlated over time in some economic applications. For variables in  $\mathbf{x}_{it}$  that are measured exactly, the corresponding elements of  $\mathbf{u}_{it}$  are zero. The parameters to be estimated consist of  $\beta$ ,  $\Phi$ , the variance-covariance parameters for  $e_{it}$  and  $\mathbf{u}_{it}$ , and the parameters associated with the distribution of the true values  $\mathbf{x}_{it}$ .

Recall that the standard errors-in-variables regression model cannot be identified without some information on measurement error variances or without instrumental variables. In applications in the social sciences, no information on the magnitude of the measurement error variances is available. However, the random coefficient errors-in-variables model (1)-(2) can be identified without any additional information, despite the fact that the random coefficient covariance matrix  $\Phi$  also needs to be estimated.

An intuitive reason is that the panel data structure indirectly provides some type of instrumental variables (see Griliches and Hausman (1986) and Paterno and Amemiya (1997)). This identification issue is discussed in the next section.

The literature on the random coefficient model with no measurement error is extensive. The basic estimation issues are discussed in Swamy (1970), Carter and Yang (1986), Harville (1977), Laird and Ware (1982), and Gumpertz and Pantula (1989). Applied work using the model without measurement error includes Nerlove (1965) for economics, Goldstein (1979, 1986) for education data, and Carter *et al.* (1986) for biometrics. Very few studies have examined measurement error models for panel data. Griliches and Hausman (1986), Wansbeek and Koning (1989), and Paterno and Amemiya (1997) discussed identification and estimation of the standard random effect model, a special case of the random coefficient model, when all or some explanatory variables are measured with error. To our knowledge, the random coefficient model (1) with measurement error structure (2) has not been previously examined.

This paper deals with a very general form of the random coefficient model with errors-in-variables given by (1)-(2). The approach of Griliches and Hausman (1986) for the random effect model cannot be directly extended to the random coefficient problem. Our approach here is related to that of Paterno and Amemiya (1997) for the random effect model. In this paper, without specifying the distributional form of  $\beta_i$ ,  $\mathbf{x}_{it}$ ,  $e_{it}$ , and  $\mathbf{u}_{it}$ , conditions for identification are derived, and estimators are proposed. Large sample properties of the estimators and results of a simulation study are also given.

## Model and Identification

To represent the general random coefficient model with some non-random coefficient, we assume that  $T \times k$  matrix of explanatory variables (excluding the intercept) can be divided into four submatrices of order  $T \times k_\ell$  ( $\ell = 1, 2, 3, 4$ ,  $k = k_1 + k_2 + k_3 + k_4$ ),

$$\mathbf{x}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i}, \mathbf{x}_{3i}, \mathbf{x}_{4i}).$$

We assume that the coefficients of  $(\mathbf{x}_{1i}, \mathbf{x}_{2i})$  are considered random over individuals, and that the coefficients of  $(\mathbf{x}_{3i}, \mathbf{x}_{4i})$  are common for all individuals. In addition,  $\mathbf{x}_{1i}$  and  $\mathbf{x}_{3i}$  are measured exactly, and  $\mathbf{x}_{2i}$  and  $\mathbf{x}_{4i}$  are measured with error.

Then, a general random coefficient errors-in-variables model for the  $i$ -th individual's  $T \times 1$  response  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})'$  and  $T \times k$  observed explanatory variable  $\mathbf{X}_i$  ( $i = 1, 2, \dots, n$ ) is

$$\mathbf{Y}_i = \mathbf{1}_T \beta_{0i} + \mathbf{x}_{1i} \beta_{1i} + \mathbf{x}_{2i} \beta_{2i} + \mathbf{x}_{3i} \beta_3 + \mathbf{x}_{4i} \beta_4 + \mathbf{e}_i, \quad (3a)$$

$$\mathbf{X}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i}, \mathbf{x}_{3i}, \mathbf{x}_{4i}) + (0, \mathbf{u}_{2i}, 0, \mathbf{u}_{4i}), \quad (3b)$$



$$E[\mathbf{e}_i] = \mathbf{0}, \quad V[\mathbf{e}_i] = \sigma_{ee} \mathbf{I}_T. \quad (3c)$$

$$E \left[ \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} \right] = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad V \left[ \begin{pmatrix} \beta_0 \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} \right] = \Phi, \quad (3d)$$

$$E[\mathbf{u}_{2i}] = \mathbf{0}, \quad V[\text{vec } \mathbf{u}_{2i}] = \Sigma_{uu_{22}} = \begin{bmatrix} \Sigma_{uu_{22}}^{(1)} & 0 & \dots & 0 \\ 0 & \Sigma_{uu_{22}}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{uu_{22}}^{(k_2)} \end{bmatrix}. \quad (3e)$$

$$E[\mathbf{u}_{4i}] = \mathbf{0}, \quad V[\text{vec } \mathbf{u}_{4i}] = \Sigma_{uu_{44}} = \begin{bmatrix} \Sigma_{uu_{44}}^{(1)} & 0 & \dots & 0 \\ 0 & \Sigma_{uu_{44}}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{uu_{44}}^{(k_4)} \end{bmatrix}. \quad (3f)$$

$$E[\mathbf{x}_i] = \boldsymbol{\mu}_x, \quad V[\text{vec } \mathbf{x}_i] = \Sigma_{xx}, \quad (3g)$$

where  $\mathbf{x}_i$ ,  $(\beta_{0i}, \beta'_{1i}, \beta'_{2i})'$ ,  $\mathbf{e}_i$ ,  $\mathbf{u}_{2i}$ , and  $\mathbf{u}_{4i}$  are independent, and  $\text{vec } (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m) = (\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_m)'$ .

Let  $\boldsymbol{\theta}$  be a vector parameter consisting of  $(\beta_0, \beta'_1, \beta'_2, \beta'_3, \beta'_4)'$ ,  $\text{vech } \Phi$ ,  $\sigma_{ee}$ ,  $\text{vech } \Sigma_{uu_{22}}^{(m)}$ ,  $m = 1, 2, \dots, k_2$ ,  $\text{vech } \Sigma_{uu_{44}}^{(m)}$ ,  $m = 1, 2, \dots, k_4$ ,  $\boldsymbol{\mu}_x$ , and  $\text{vech } \Sigma_{xx}$ , where  $\text{vech } \mathbf{A}$  denotes the  $\frac{p(p+1)}{2} \times 1$  vector created by listing the elements of a  $p \times p$  matrix  $\mathbf{A}$  on and below its diagonal starting with the first column. The parameter vector  $\boldsymbol{\theta}$  needs to be identified based on the observations

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_i \\ \text{vec } \mathbf{X}_i \end{pmatrix}, \quad i = 1, 2, \dots, n.$$

Note that model (3) does not specify any distributional form of  $\beta_{0i}$ ,  $\beta_{1i}$ ,  $\beta_{2i}$ ,  $\mathbf{e}_i$ ,  $\mathbf{u}_{2i}$ ,  $\mathbf{u}_{4i}$ , and  $\mathbf{x}_i$ , except for the first two moments. Thus, it may be natural to consider the identification of  $\boldsymbol{\theta}$  only through the first two moments of  $\mathbf{Z}_i$ ,

$$\boldsymbol{\mu}_Z(\boldsymbol{\theta}) = E[\mathbf{Z}_i], \quad (4a)$$

$$\Sigma_{ZZ}(\boldsymbol{\theta}) = V[\mathbf{Z}_i]. \quad (4b)$$

We investigate whether or not  $\boldsymbol{\theta}$  can be uniquely determined given  $\boldsymbol{\mu}_Z(\boldsymbol{\theta})$  and  $\Sigma_{ZZ}(\boldsymbol{\theta})$ . This approach falls under the general method known as structural equation modeling or moment structure analysis that is widely used in social and behavioral sciences (see, e.g., Bollen (1989), Bollen and Long (1993), and Hoyle (1995)).

A simple sufficient condition for identification of the (mean) regression coefficient  $\boldsymbol{\beta} = (\beta_0, \beta'_1, \beta'_2, \beta'_3, \beta'_4)'$  is given in the first Theorem.

**Theorem 1**  $\beta$  can be uniquely determined from  $\mu_Z(\theta)$  in (4a) if

$$\text{rank} \begin{bmatrix} \mathbf{1}_T & \mu_x \end{bmatrix} = k + 1. \quad (5)$$

**Proof:** This is immediate since

$$\mu_Z(\theta) = \begin{pmatrix} \begin{bmatrix} \mathbf{1}_T & \mu_x \end{bmatrix} \beta \\ \text{vec } \mu_x \end{pmatrix}. \quad \square$$

Recall that  $\mu_x$  is  $T \times k$ . Thus, if  $T > k$  and if the (true)  $x$ -variables have sufficient variability on average, then  $\beta$  can be identified. Up to this point,  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  has been treated as a random sample. If  $\mathbf{x}_i$  are considered fixed constants, then condition (5) with  $\mu_x = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  is sufficient for the existence of a consistent estimator of  $\beta$ .

The following theorem gives conditions for the identification of the whole  $\theta$ . For this, let

$$\begin{aligned} \mathbf{x}_i^{(R)} &= (\mathbf{1}_T, \mathbf{x}_{1i}, \mathbf{x}_{2i}) = \begin{pmatrix} \mathbf{v}'_{1i} \\ \vdots \\ \mathbf{v}'_{Ti} \end{pmatrix}, \\ C_{j\ell} &= E[\mathbf{v}_{\ell i} \mathbf{v}'_{ji}], \quad j, \ell = 1, 2, \dots, T, \end{aligned}$$

and let  $\mathbf{K}_p$  denote the matrix of zeros and ones such that  $\text{vec } \mathbf{A} = \mathbf{K}_p \text{vech } \mathbf{A}$  for any  $p \times p$  symmetric  $\mathbf{A}$ .

**Theorem 2**  $\theta$  can be uniquely determined from  $\mu_Z(\theta)$  and  $\Sigma_{ZZ}(\theta)$  in (4) if (5) holds, if no element of  $\beta_2$  and  $\beta_4$  is zero, and if the  $\frac{T(T+1)}{2} \times [\frac{(k_1+k_2+1)(k_1+k_2+2)}{2} + 1]$  matrix

$$\left[ \begin{pmatrix} (\text{vec } C_{11})' \\ \vdots \\ (\text{vec } C_{T1})' \\ (\text{vec } C_{22})' \\ \vdots \\ (\text{vec } C_{T2})' \\ \vdots \\ (\text{vec } C_{TT})' \end{pmatrix} \quad \mathbf{K}_{(k_1+k_2+1)}, \quad \text{vech } \mathbf{I}_T \right] \quad (6)$$

has full column rank.

**Proof:** With (5),  $\mu_x$  and  $\beta$  are determined from  $\mu_Z(\theta)$ . Given  $\beta$ , consider a transformation

$$\mathbf{Y}_i^* = \mathbf{Y}_i - (\mathbf{1}_T \quad \mathbf{X}_i) \beta = (\mathbf{1}_T \quad \mathbf{x}_{1i} \quad \mathbf{x}_{2i}) \begin{pmatrix} \beta_{0i} - \beta_0 \\ \beta_{1i} - \beta_1 \\ \beta_{2i} - \beta_2 \end{pmatrix} - (\mathbf{u}_{2i} \quad \mathbf{u}_{4i}) \begin{pmatrix} \beta_2 \\ \beta_4 \end{pmatrix} + \mathbf{e}_i.$$

Then,

$$\begin{aligned} V[\mathbf{Y}_i^*] &= \Upsilon + (\beta_2' \otimes \mathbf{I}_T) \Sigma_{uu_{22}} (\beta_2 \otimes \mathbf{I}_T) + (\beta_4' \otimes \mathbf{I}_T) \Sigma_{uu_{44}} (\beta_4 \otimes \mathbf{I}_T) + \sigma_{ee} \mathbf{I}_T, \\ \Upsilon &= \{\Upsilon_{j\ell}\}, \\ \Upsilon_{j\ell} &= \text{tr}(\mathbf{C}_{j\ell} \Phi), \\ \text{Cov}[\mathbf{Y}_i^*, \text{vec } \mathbf{X}_i] &= [0, -\beta_{21} \Sigma_{uu_{22}}^{(1)}, \dots, -\beta_{2k_2} \Sigma_{uu_{22}}^{(k_2)}, 0, -\beta_{41} \Sigma_{uu_{44}}^{(1)}, \dots, -\beta_{4k_4} \Sigma_{uu_{44}}^{(k_4)}], \\ V[\text{vec } \mathbf{X}_i] &= \Sigma_{xx} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_{uu_{22}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Sigma_{uu_{44}} \end{bmatrix}. \end{aligned}$$

where  $\beta_2 = (\beta_{21}, \dots, \beta_{2k_2})'$  and  $\beta_4 = (\beta_{41}, \dots, \beta_{4k_4})'$ . With the nonzero condition on  $\beta_2$  and  $\beta_4$ ,  $\Sigma_{uu_{22}}$  and  $\Sigma_{uu_{44}}$  are determined from  $\text{Cov}[\mathbf{Y}_i^*, \text{vec } \mathbf{X}_i]$ . Thus,  $\Sigma_{xx}$  is determined from  $V[\text{vec } \mathbf{X}_i]$ . Given  $\beta$ ,  $\Sigma_{uu_{22}}$ , and  $\Sigma_{uu_{44}}$ ,  $\Phi$  and  $\sigma_{ee}$  can be uniquely obtained from  $V[\mathbf{Y}_i^*]$  under the full column rank condition in the theorem.  $\square$

Note that  $\mathbf{C}_{j\ell}$ ,  $j, \ell = 1, 2, \dots, T$ , represent the changes in the second moments of the elements of  $(\mathbf{1}_T, \mathbf{x}_{1i}, \mathbf{x}_{2i})$  over time. Thus, if the true  $x$ -variables corresponding to the random coefficients vary sufficiently over time to the extent of nonconstant second moments, then all parameters in model (3) can be identified from the first two moments of  $\mathbf{Z}_i$ . The conditions in Theorem 1 and 2 mean intuitively that the (true)  $x$ -variables cannot have redundancy over each other and time in the first two moments. A necessary condition is that  $T$  is large enough compared to  $k_\ell$ ,  $\ell = 1, 2, 3, 4$ , i.e.,

$$\begin{aligned} T &\geq k + 1, \text{ and} \\ \frac{T(T+1)}{2} &\geq \frac{(k_1 + k_2 + 1)(k_1 + k_2 + 2)}{2} + 1. \end{aligned}$$

Hence, if  $T$  is large enough and if the true  $x$ -values differ sufficiently, then all parameters in the random coefficient errors-in-variables model (3) are identified without requiring the availability of any additional information.

Throughout, the measurement error for each explanatory variable measured with error is assumed to have any unspecified covariance matrix over  $T$  time points. Since such a general error covariance matrix

can be identified, any more structured measurement error covariance matrix can also be identified under the same conditions.

## Estimation

As in the identification discussion, we consider estimation of all parameters in model (3) without specifying the distributional form of  $(\beta_{0i}, \beta'_{1i}, \beta'_{2i})'$ ,  $e_i$ ,  $u_{2i}$ ,  $u_{4i}$ , and  $x_i$ . A natural approach to this is to base our estimation on the first two sample moments. Let

$$\begin{aligned} \mathbf{W}_i &= \begin{pmatrix} \mathbf{Z}_i \\ \text{vech}(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})' \end{pmatrix}, \\ \bar{\mathbf{Z}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i, \\ \bar{\mathbf{W}} &= \begin{pmatrix} \bar{\mathbf{Z}} \\ \text{vech } \mathbf{S} \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i, \\ \mathbf{S} &= \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})'. \end{aligned}$$

Under the identification condition in the previous section, a method of moments estimator of  $\theta$  can be obtained by minimizing some distance between  $\bar{\mathbf{W}}$  and  $\eta(\theta)$ . A natural measure for distance without specifying the distribution of  $\mathbf{W}_i$  is

$$Q(\theta) = n[\bar{\mathbf{W}} - \eta(\theta)]' \hat{\Omega}^{-1} [\bar{\mathbf{W}} - \eta(\theta)], \quad (7)$$

where  $\hat{\Omega}$  is a distribution-free estimator of  $V[\mathbf{W}_i]$  given by

$$\hat{\Omega} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{W}_i - \bar{\mathbf{W}})(\mathbf{W}_i - \bar{\mathbf{W}})'. \quad (8)$$

We call the value of  $\theta$  minimizing  $Q(\theta)$  in (7) the generalized least squares (GLS) estimator of  $\theta$  and denote it by  $\hat{\theta}_{\text{GLS}}$ .

Studies in the covariance structure analysis literature (e.g., Hu *et al.*, 1992 and West *et al.*, 1995) have pointed out that  $\hat{\Omega}$  in (8) tends to have a large variance even in fairly large samples, making  $\hat{\theta}_{\text{GLS}}$  unstable and variable. Thus, despite the large sample optimality of  $\hat{\theta}_{\text{GLS}}$ , we consider an alternative estimator that may be less variable in finite samples because of its simplicity. Let

$$\begin{aligned} L(\theta) &= n\{[\bar{\mathbf{Z}} - \mu_Z(\theta)]' \Sigma_{ZZ}^{-1}(\theta) [\bar{\mathbf{Z}} - \mu_Z(\theta)] + \ln |\Sigma_{ZZ}(\theta)| \\ &\quad + \text{tr} [\mathbf{S} \Sigma_{ZZ}^{-1}(\theta)] - \ln |\mathbf{S}| - T(k+1)\}. \end{aligned} \quad (9)$$

If  $\mathbf{Z}_i$  is normally distributed, the value of  $\theta$  minimizing  $L(\theta)$  is the maximum likelihood estimator. In the random coefficient models, the observation  $\mathbf{Z}_i$  is not normally distributed. Considering  $L(\theta)$  as a distance between  $\bar{\mathbf{W}}$  and  $\eta(\theta)$ , the pseudo maximum likelihood (PML) estimator, denoted by  $\hat{\theta}_{PML}$  and obtained by minimizing (9), is expected to be a reasonable estimator without specifying the distributional form. Note that  $\hat{\theta}_{PML}$  does not involve any statistic other than  $\bar{\mathbf{W}}$  and that  $\hat{\theta}_{PML}$  can be conveniently computed using standard packages for moment structure analysis or structural equation modeling (e.g., PROC CALIS (SAS Institute, 1989)).

The following theorem gives the asymptotic distributions of  $\hat{\theta}_{GLS}$  and  $\hat{\theta}_{PML}$  under a very general condition. The large sample result applies when the number of individuals  $n$  is large while the number of time points  $T$  may not be large. Let  $\theta_0$  denote the true value of  $\theta$ .

**Theorem 3** *Assume a general identification condition*

(a) *for any  $\epsilon > 0$  there is a  $\delta > 0$  such that if  $|\theta - \theta_0| > \epsilon$  then  $|\eta(\theta) - \eta(\theta_0)| > \delta$ .*

*Also assume that the distribution of  $\mathbf{e}_i$ ,  $\mathbf{u}_{2i}$ ,  $\mathbf{u}_{4i}$ ,  $(\beta_0, \beta'_{1i}, \beta'_{2i})'$  and  $\mathbf{x}_i$  is such that either*

*(b - i)  $\mathbf{Z}_i$ 's are independent and identically distributed with finite fourth moments,*

*or*

*(b - ii)  $\mathbf{Z}_i$ 's are independent and possess bounded  $4 + \gamma$  moments for some  $\gamma > 0$ .*

*Then, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\theta}_{GLS} - \theta_0) \xrightarrow{L} N(0, \Gamma_{GLS}),$$

$$\sqrt{n}(\hat{\theta}_{PML} - \theta_0) \xrightarrow{L} N(0, \Gamma_{PML}),$$

*where*

$$\begin{aligned} \Gamma_{GLS} &= (\mathbf{F}'_0 \Omega_0^{-1} \mathbf{F}_0)^{-1}, \\ \Gamma_{PML} &= (\mathbf{F}'_0 \Psi_0^{-1} \mathbf{F}_0)^{-1} \mathbf{F}'_0 \Psi_0^{-1} \Omega_0 \Psi_0^{-1} \mathbf{F}_0 (\mathbf{F}'_0 \Psi_0^{-1} \mathbf{F}_0)^{-1}, \\ \mathbf{F}_0 = \mathbf{F}(\theta_0) &= \frac{\partial \eta(\theta_0)}{\partial \theta'}, \\ \Omega_0 &= \text{plim } \hat{\Omega}, \\ \Psi_0 = \Psi(\theta_0) &= \begin{bmatrix} \Sigma_{ZZ}(\theta_0) & 0 \\ 0 & 2\mathbf{K}^+ [\Sigma_{ZZ}(\theta_0) \otimes \Sigma_{ZZ}(\theta_0)] \mathbf{K}^{+'} \end{bmatrix}, \end{aligned}$$

$\hat{\Omega}$  is given in (8),  $\mathbf{K}^+ = (\mathbf{K}'_{k+1} \mathbf{K}_{k+1})^{-1} \mathbf{K}_{k+1}$ , and  $\mathbf{K}_p$  is defined before Theorem 2.

**Proof:** The consistency of  $\hat{\theta}_{GLS}$  and  $\hat{\theta}_{PML}$  follows from the identification condition (a), the form of the two distance measures, and the fact that  $\text{plim } \bar{\mathbf{W}} = \eta(\theta_0)$  under (b - i) or (b - ii). The asymptotic normality is a consequence of the asymptotic normality of  $\bar{\mathbf{W}}$  under (b - i) or (b - ii) and of the derivative

form of  $Q(\theta)$  and  $L(\theta)$ . The limiting covariance forms can be derived using the standard derivative equations. The lower right-hand corner of  $\Psi_0$  is the limiting covariance matrix of  $\text{vech } S$  under the normality of  $Z_i$ .  $\square$

Thus, estimated asymptotic covariance matrices of  $\hat{\theta}_{\text{GLS}}$  and  $\hat{\theta}_{\text{PML}}$  can be obtained as

$$\hat{V}[\hat{\theta}_{\text{GLS}}] = \frac{1}{n} \left[ \mathbf{F}'(\hat{\theta}_{\text{GLS}}) \hat{\Omega}^{-1} \mathbf{F}(\hat{\theta}_{\text{GLS}}) \right]^{-1}, \quad (10)$$

$$\begin{aligned} \hat{V}[\hat{\theta}_{\text{PML}}] &= \frac{1}{n} \left[ \mathbf{F}'(\hat{\theta}_{\text{PML}}) \Psi(\hat{\theta}_{\text{PML}})^{-1} \mathbf{F}(\hat{\theta}_{\text{PML}}) \right]^{-1} \mathbf{F}'(\hat{\theta}_{\text{PML}}) \Psi(\hat{\theta}_{\text{PML}})^{-1} \Omega^{-1} \Psi^{-1} \mathbf{F}(\hat{\theta}_{\text{PML}}) \\ &\quad \left[ \mathbf{F}'(\hat{\theta}_{\text{PML}}) \Psi(\hat{\theta}_{\text{PML}})^{-1} \mathbf{F}(\hat{\theta}_{\text{PML}}) \right]^{-1}. \end{aligned} \quad (11)$$

which can be used for asymptotic inferences for  $\theta$ . In a very large sample, we would expect the inferences based on  $\hat{\theta}_{\text{GLS}}$  and  $\hat{V}[\hat{\Gamma}_{\text{GLS}}]$  to be more efficient than those using  $\hat{\theta}_{\text{PML}}$  and  $\hat{V}[\hat{\Gamma}_{\text{PML}}]$ . However, the computation of  $\hat{\theta}_{\text{PML}}$  does not involve any sample higher order moments in minimization, and  $\hat{\Omega}$  appears only in the evaluation of  $\hat{V}[\hat{\Gamma}_{\text{PML}}]$ . Thus, in small to moderate samples, the inference procedure based on  $\hat{\theta}_{\text{GLS}}$  and  $\hat{V}[\hat{\Gamma}_{\text{GLS}}]$  may not be as efficient or accurate as those based on  $\hat{\theta}_{\text{PML}}$  and  $\hat{V}[\hat{\Gamma}_{\text{PML}}]$  because of the high variability in  $\hat{\Omega}$ . These issues are addressed in the next section using a simulation study.

## Simulation Study

To examine the behavior of the GLS and the PML estimators in finite samples, 1000 samples of size  $n = 50, 200$ , and  $500$  were generated according to model (3) with  $k_1 = k_3 = k_4 = 0$ ,  $k_2 = 1$ , and  $T = 3$ . We write the model as

$$\begin{aligned} \mathbf{Y}_i &= \begin{pmatrix} 1_3 & \mathbf{X}_i \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + \mathbf{e}_i, \\ \mathbf{X}_i &= \mathbf{x}_i + \mathbf{u}_i, \\ E \left[ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \right] &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \\ \Phi &= \begin{bmatrix} \phi_{\alpha\alpha} & \phi_{\beta\alpha} \\ \phi_{\beta\alpha} & \phi_{\beta\beta} \end{bmatrix}. \end{aligned}$$

Stochastic components in the model,  $(\alpha_i, \beta_i)$ ,  $\mathbf{x}_i$ ,  $\mathbf{u}_i$  and  $\mathbf{e}_i$ , were generated as normally distributed variates. Note that this data generation procedure results in normally distributed  $\mathbf{X}_i$ 's but not normally distributed  $\mathbf{Y}_i$ 's. Thus, the observed vector  $\mathbf{Z}_i$  is not normally distributed. For simplicity, the measurement errors  $e_{it}$  and  $u_{it}$  were generated as independently and identically distributed over time

and over individuals, and thus  $\text{var}[e_i] = \sigma_{ee}\mathbf{I}_3$  and  $\text{var}[u_i] = \sigma_{uu}\mathbf{I}_3$ . The true parameter values were set to be

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \quad \Phi = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \sigma_{ee} = 4, \quad \sigma_{uu} = 4,$$

$$\mu_x = \begin{pmatrix} 2 \\ 5 \\ 10 \end{pmatrix}, \quad \text{and} \quad \Sigma_{xx} = \begin{bmatrix} 4.0 & 2.0 & 0.8 \\ 2.0 & 4.8 & 2.8 \\ 0.8 & 2.8 & 5.6 \end{bmatrix}.$$

Model (3) with  $e_{it}$  and  $u_{it}$  treated as independent and identically distributed over time and over individuals was fitted to each sample via generalized least squares and via pseudo maximum likelihood. The GLS estimator,  $\hat{\theta}_{\text{GLS}}$ , was obtained by minimizing (7) while the PML estimator,  $\hat{\theta}_{\text{PML}}$ , was obtained by minimizing (9). In both minimization algorithms,  $\Phi$  was re-parameterized to be the lower triangular  $\Gamma$  satisfying  $\Phi = \Gamma\Gamma'$ , so that the resulting estimate of  $\Phi$  is always in the parameter space.

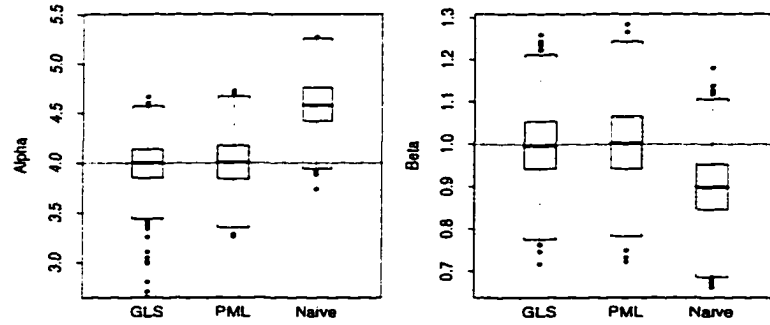


Figure 1 Box plots of estimates of  $\alpha$  and  $\beta$  ( $n = 500$ )

Figure 1 gives box plots of estimates of  $\beta$  for  $n = 500$ , including a naive estimator that is produced by standard random coefficient analysis ignoring the existence of measurement error. In Figure 1, there are 17 observations of  $\hat{\mu}_{\alpha, \text{GLS}}$  that are less than 2.6 and have been omitted from the box plot of estimates of  $\mu_{\alpha}$ . In these and all other box plots, the true value of the parameter is indicated by a dotted horizontal line. Figure 1 clearly shows that the naive estimator ignoring measurement error is useless due to its large bias. On the other hand, the PML and the GLS estimators have very small bias, and the major portions of their sampling distributions are roughly symmetric around the true value.

Figure 2 presents box plots of the GLS and the PML estimators of  $\alpha$  and  $\beta$ , where the scales of the plots are kept constant over  $n = 50, 200$ , and  $500$ . For estimating  $\alpha$ , the GLS estimator tends to

take values very far away from the true value, possesses some bias, and has a heavy-tailed sampling distribution even with large  $n$ . The PML estimator of  $\alpha$  is more stable and has a sampling distribution that is more tightly concentrated around the true value. For  $\beta$ , the GLS estimator's high variability starts to disappear with  $n = 200$ . However, despite the inferiority in the limiting distribution, the PML estimator of  $\beta$  is as efficient as the GLS estimator even for  $n = 500$ .

Figure 3 gives box plots of the GLS and PML estimators of the elements  $\phi_{\alpha\alpha}$ ,  $\phi_{\beta\alpha}$ , and  $\phi_{\beta\beta}$  of the random coefficient covariance matrix  $\Phi$ . For  $\phi_{\alpha\alpha}$  and  $\phi_{\beta\alpha}$ , the two estimators' sampling distributions are rather similar to each other for all  $n$ , but for estimating  $\phi_{\beta\beta}$ , the GLS estimator produced a large number of outlying values. Figure 4 presents box plots for estimating some elements of  $\mu_x$  and  $\Sigma_{xx}$ . For  $\mu_x$ , the two estimators were very similar to each other. For the elements of  $\Sigma_{xx}$ , the GLS estimator produced a large number of extreme values and was more variable than the PML estimator. As shown in Figure 5, for the error variances  $\sigma_{ee}$  and  $\sigma_{uu}$ , the PML estimator was much more stable and less variable than the GLS estimator. Therefore, for samples of sizes considered here, the PML estimator may be considered superior to the GLS estimator in terms of finite sample stability and variability.

The other basis for comparison is how well the estimators lead to valid inferences about the parameters. Table 1 lists the simulation variances of the estimates and the average of the estimated variances using (10) and (11) for the case with  $n = 500$ . The estimated covariance matrix (11) for PML is a good estimator of the variability of the PML estimator in the sense that the average estimated variances for the PML were close to the Monte Carlo variances. For GLS, the estimated variances using (10) are not as close to the simulation variances. For estimating  $\Phi$ , some very large variance estimates seem to dominate the variance estimates. For other parameters, the estimated variances generally underestimate the true variability.

Table 2 gives the simulated coverage probabilities of nominal 95% confidence intervals formed using the GLS and the PML estimators and their estimated standard errors using (10) and (11), respectively. In small samples, the GLS coverage probabilities are unsatisfactorily less than the nominal level whereas the PML coverage probabilities were not very different from the nominal level of 0.95. In samples of moderate size, both the GLS and the PML methods generated confidence intervals with coverage probabilities that were similar to 0.95. However, especially for the variance-covariance parameters, the GLS coverage probabilities are less than the nominal level. The PML approach produced confidence intervals with coverage probabilities closer to the nominal level than the GLS approach. Although not reported fully here, we conducted similar studies where the  $x_i$ 's were generated as chi-square random variables and separately, as fixed constants over repeated samples. The coverage probabilities for the



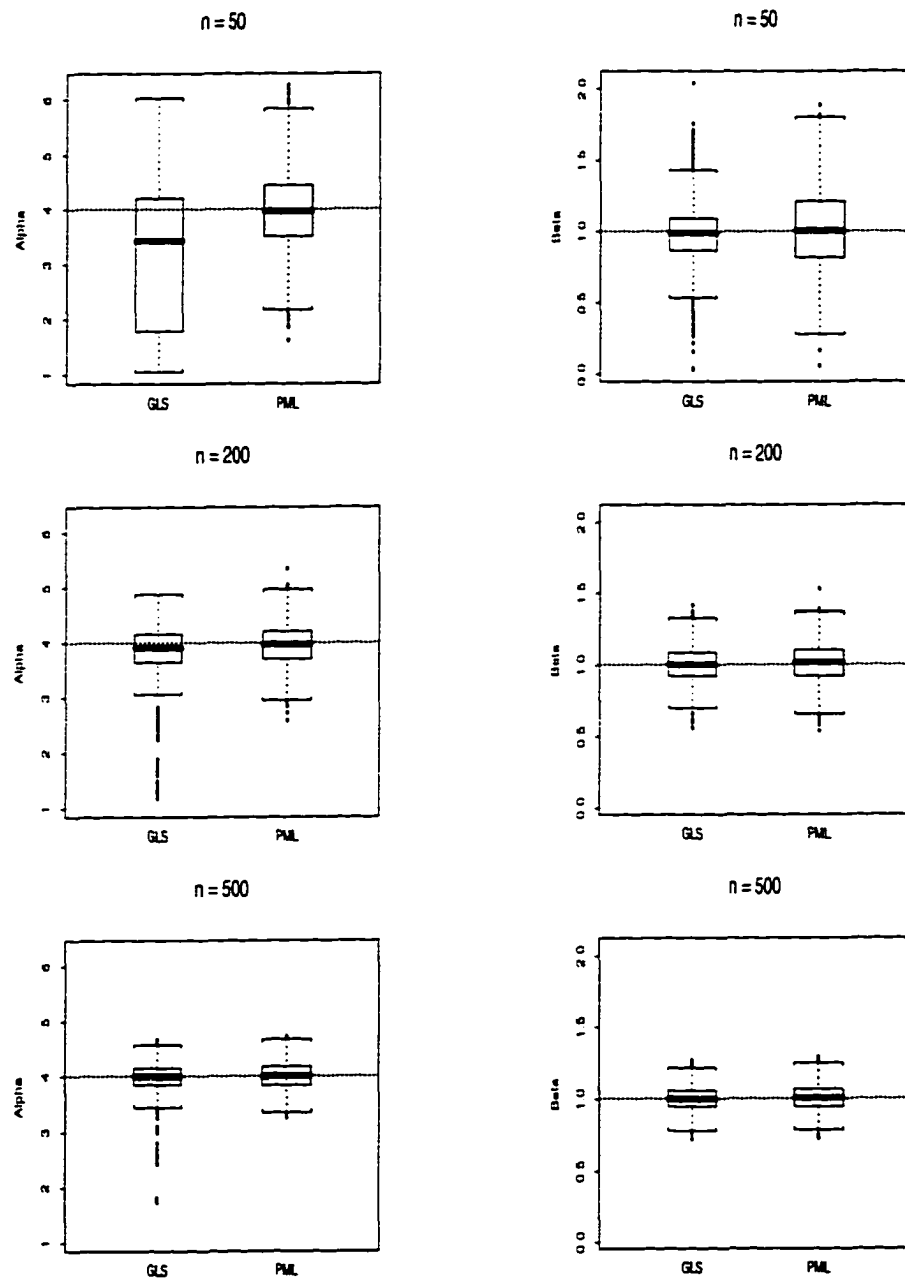
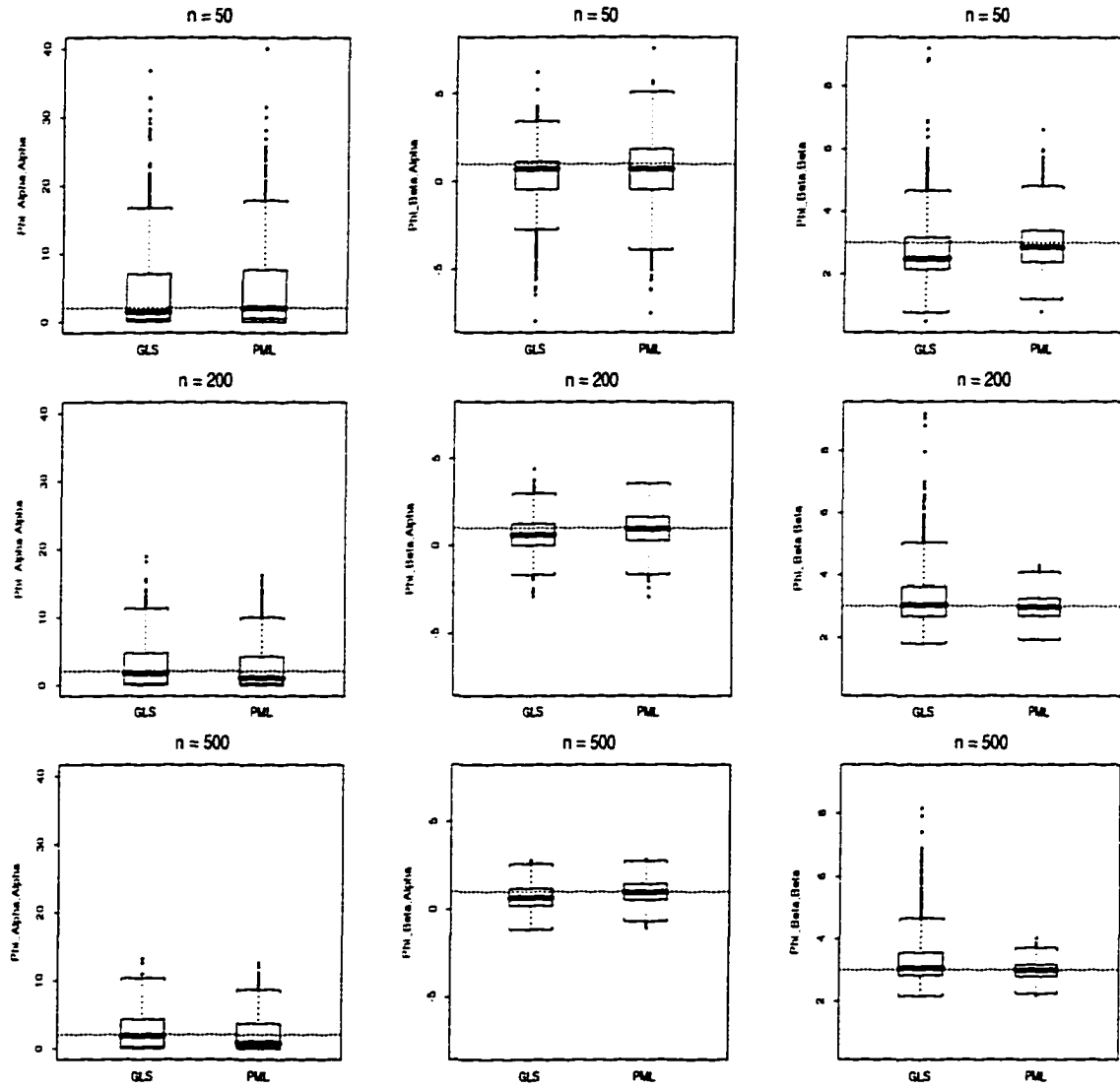


Figure 2 Box plots of estimates of  $\alpha$  and  $\beta$

Figure 3 Box plots of estimates of  $\Phi$

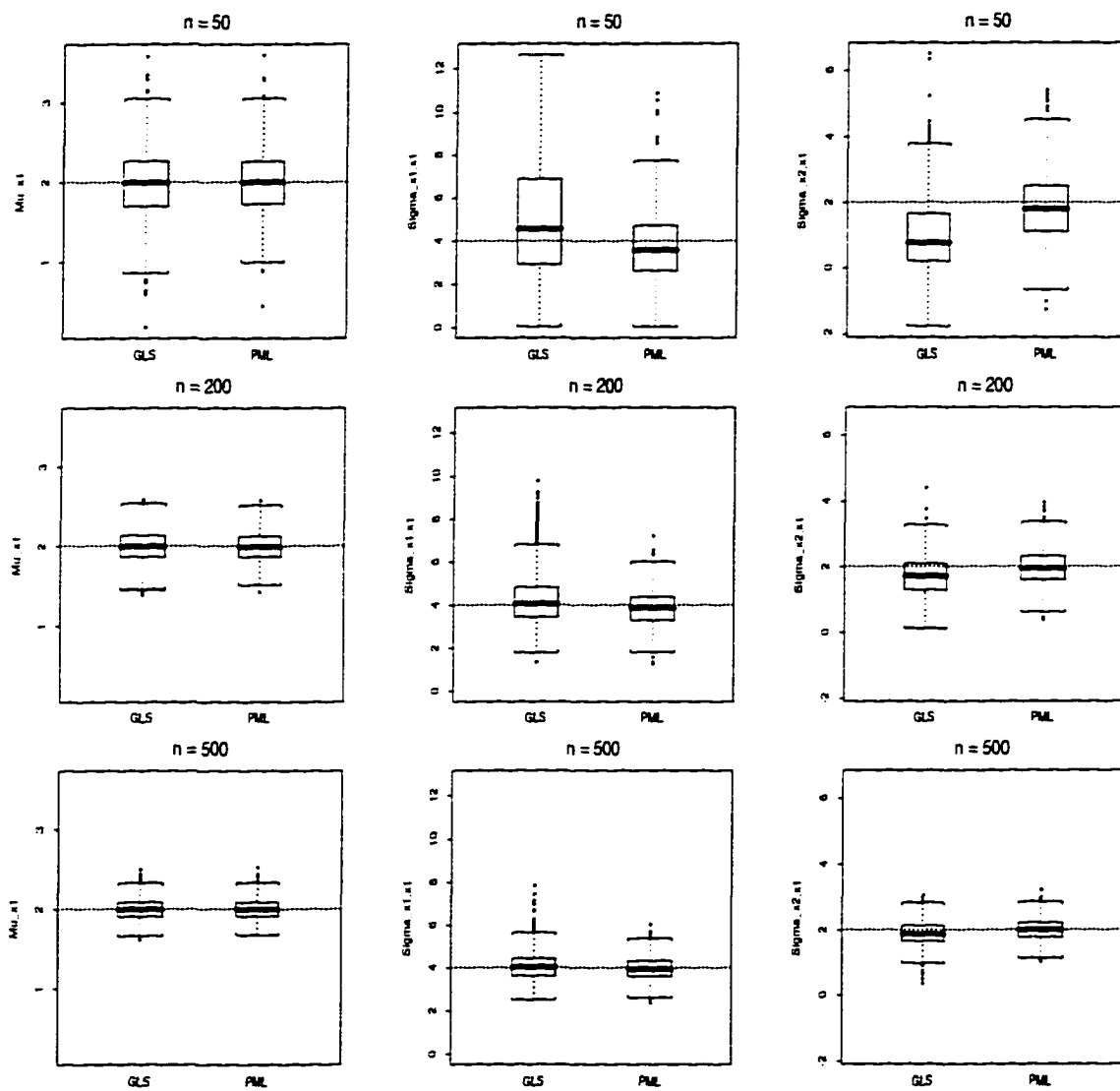


Figure 4 Box plots of estimates of  $\mu_x$ ,  $\Sigma_{x_1,x_1}$  and  $\Sigma_{x_2,x_1}$

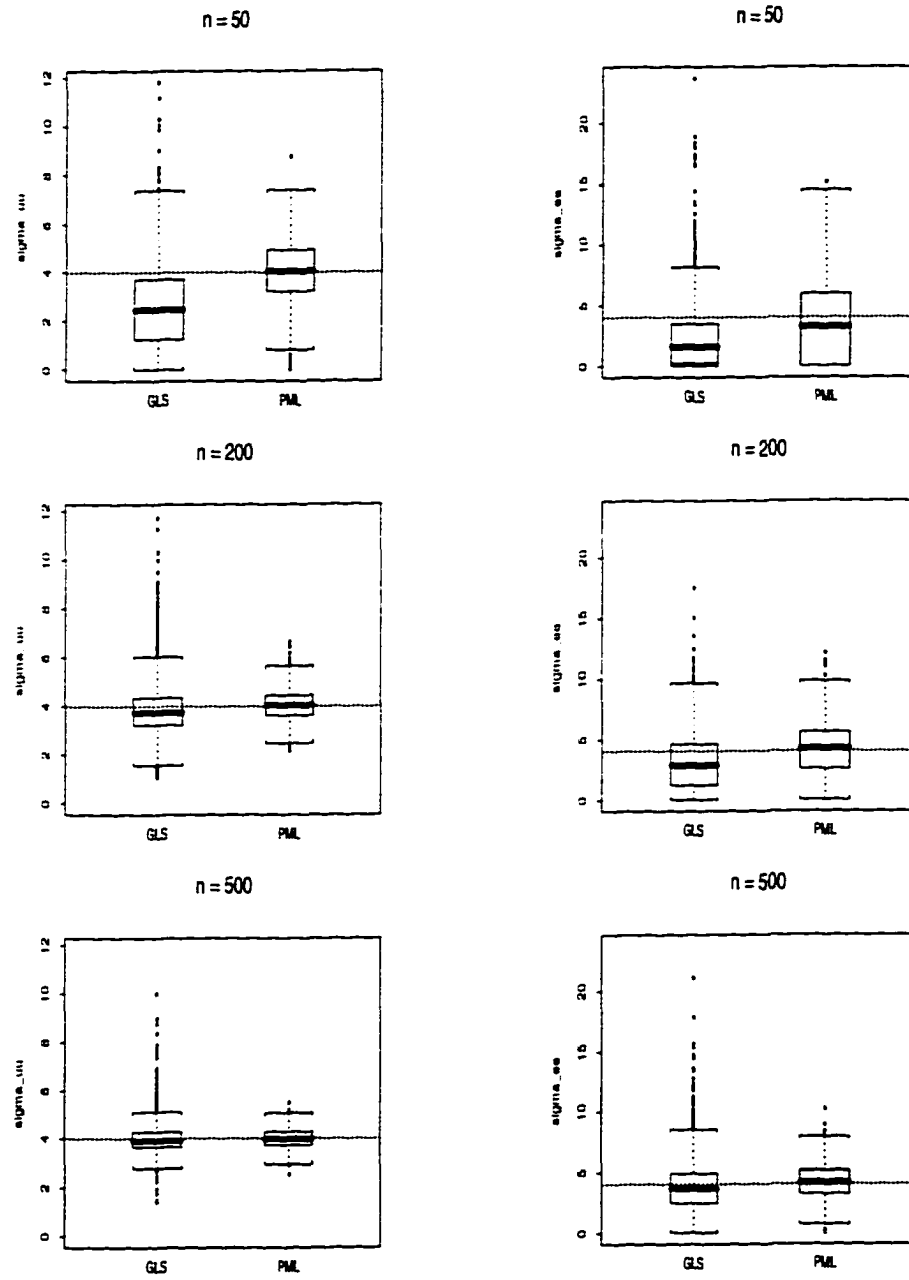


Figure 5 Box plots of estimates of  $\sigma_{uu}$  and  $\sigma_{ee}$

Table 1 Variance estimates for parameters ( $n = 500$ )

Parameter	GLS		PML	
	Simulation variance	Average estimated variance <sup>a</sup>	Simulation variance	Average estimated variance
$\alpha$	0.0957	0.0396	0.0626	0.0583
$\beta$	0.0070	0.0066	0.0077	0.0079
$\phi_{\alpha\alpha}$	6.0938	$3.92 \times 10^9$	6.1081	9.7686
$\phi_{\beta\alpha}$	0.4640	$1.73 \times 10^{11}$	0.4057	0.4520
$\phi_{\beta\beta}$	0.6265	$1.57 \times 10^{10}$	0.0735	0.0668
$\mu_{x_1}$	0.0161	0.0144	0.0159	0.0151
$\mu_{x_2}$	0.0162	0.0140	0.0163	0.0159
$\mu_{x_3}$	0.0189	0.0182	0.0183	0.0187
$\sigma_{x_1, x_1}$	0.4587	0.3062	0.3027	0.2720
$\sigma_{x_2, x_1}$	0.1398	0.1242	0.1168	0.1208
$\sigma_{x_3, x_1}$	0.1422	0.1408	0.1406	0.1349
$\sigma_{x_2, x_2}$	0.5201	0.3537	0.3665	0.3216
$\sigma_{x_3, x_2}$	0.2366	0.1978	0.1729	0.1664
$\sigma_{x_3, x_3}$	0.7556	0.3838	0.4859	0.4617
$\sigma_{uu}$	0.6712	0.1394	0.1627	0.1499
$\sigma_{ee}$	5.6144	2.5655	2.2432	2.2729

<sup>a</sup>Median GLS variance estimates for  $\phi_{\alpha\alpha}$ ,  $\phi_{\beta\alpha}$  and  $\phi_{\beta\beta}$  were 8.9272, 0.5773, and 1.4337, respectively.

GLS and PML confidence intervals for  $\alpha$  and  $\beta$  are given in Table 3. These suggest that PML generally produces more accurate intervals than GLS, and that the PML procedure seems to be reasonably robust over different types of the true values  $\mathbf{x}_i$  for making inferences about the mean of the random coefficients.

## Summary

Random coefficient errors-in-variables models for panel data have not been previously studied. This paper examined this random coefficient errors-in-variables model and discussed identification and estimation issues. Conditions for identification were derived based on the first two moments without using a specific distributional form. The model parameters are identified essentially if enough time points are included in the sample. The identification does not require any outside information. Two estimation procedures, GLS and PML, were presented without imposing particular distributional assumptions. The limiting distributions of the two estimators were derived assuming that the number of individuals tends to infinity. The finite sample properties of the estimators were assessed using a simulation study.

Table 2 Simulated coverage probabilities of nominal 95% confidence intervals  
(normal  $\mathbf{x}_i$ )

Parameter	$n = 50$		$n = 200$		$n = 500$	
	GLS	PML	GLS	PML	GLS	PML
$\alpha$	0.559	0.947	0.845	0.948	0.916	0.938
$\beta$	0.864	0.943	0.934	0.948	0.949	0.956
$\phi_{\alpha\alpha}$	0.942	0.987	0.971	0.983	0.972	0.979
$\phi_{\beta\alpha}$	0.885	0.978	0.929	0.964	0.936	0.966
$\phi_{\beta\beta}$	0.883	0.892	0.971	0.940	0.987	0.927
$\mu_{x_1}$	0.861	0.953	0.928	0.946	0.942	0.949
$\mu_{x_2}$	0.814	0.934	0.925	0.955	0.936	0.947
$\mu_{x_3}$	0.860	0.940	0.934	0.948	0.939	0.944
$\sigma_{x_1, x_1}$	0.703	0.910	0.851	0.908	0.911	0.931
$\sigma_{x_2, x_1}$	0.658	0.919	0.823	0.949	0.904	0.953
$\sigma_{x_3, x_1}$	0.943	0.943	0.949	0.939	0.951	0.943
$\sigma_{x_2, x_2}$	0.783	0.907	0.860	0.940	0.908	0.929
$\sigma_{x_3, x_2}$	0.652	0.945	0.840	0.938	0.897	0.934
$\sigma_{x_3, x_3}$	0.801	0.928	0.786	0.940	0.860	0.940
$\sigma_{uu}$	0.665	0.944	0.742	0.930	0.838	0.940
$\sigma_{ee}$	0.967	0.994	0.959	0.974	0.876	0.944

Table 3 Simulated coverage probabilities of nominal 95% confidence intervals  
(non-normal  $\mathbf{x}_i$ ;  $n = 500$ )

Parameter	$\mathbf{x}_i \sim \chi^2$		$\mathbf{x}_i$ fixed	
	GLS	PML	GLS	PML
$\alpha$	0.883	0.949	0.532	0.950
$\beta$	0.949	0.931	0.532	0.938

Despite its inferiority over the GLS in the limiting distribution, the PML estimator was superior in terms of finite sample stability and variability. The asymptotic inference procedures using the PML estimator are practically accurate for most sample sizes.

## References

- Bollen, K. A. (1989) *Structural Equations with Latent Variables*. New York: John Wiley.
- Bollen, K. A. and J. S. Long (1993) *Testing Structural Equation Models*. Thousand Oaks: Sage Publications.

- Carter, R. L. and M. C. K. Yang (1986) Large sample inference in random coefficient regression models. *Communications in Statistics, Part A-Theory and Methods* 15, 2507-2525.
- Carter, R. L., M. Feuerstein, and J. T. Love (1986) Repeated measures regression: An application to the analysis of mood and chronic pain. *Computational Statistics and Data Analysis* 4, 131-140.
- Goldstein, H. (1979) Some models for analyzing longitudinal data on educational attainment. *Journal of Royal Statistical Society Series A* 142, 407-32.
- Goldstein, H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73, 43-56.
- Griliches, Z. and J. A. Hausman (1986) Errors in variables in panel data. *Journal of Econometrics* 31, 93-118.
- Gumpertz, M. and S. G. Pantula (1989) A simple approach to inference in random coefficient models. *The American Statistician* 43, 203-210.
- Harville, D. A. (1977) Maximum likelihood approaches to variance components estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.
- Hoyle, R. (1995) *Structural Equation Models: Concepts, Issues and Applications*. Thousand Oaks: Sage Publications.
- Hu, L., P.M. Bentler, and Y. Kano (1992) Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin* 112, 351-362.
- Laird, N. M. and J. H. Ware (1982) Random-effects models for longitudinal data. *Biometrics* 38, 963-974.
- Nerlove, M. (1965) *Estimation and Identification of Cobb-Douglas Production Functions* Chicago: Rand McNally & Co.
- Paterno, E. M. and Y. Amemiya (1997) Analysis of panel data using a random effect errors-in-variables model, Preprint Series 97-16, Statistical Laboratory, Iowa State University.
- SAS Institute, Inc. (1991) *The CALIS procedure: analysis of covariance structures*. Cary, NC: Author.
- Swamy, P. A. V. B. (1970) Efficient Inference in Random Coefficient Regression Models. *Econometrica* 38, 311-323.
- Wansbeek, R. J. and R. H. Koning (1989) Measurement error and panel data. *Statistica Neerlandica* 45, 85-92.
- West, S. G., J. F. Finch and P. J. Curran (1995) Structural equation models with nonnormal data. In: R. Hoyle (ed.), *Structural Equation Models: Concepts, Issues and Applications*. Thousand Oaks: Sage Publications.

## GENERAL CONCLUSIONS

This dissertation has examined measurement error models for panel data or repeated measures. Such models have not been widely studied, yet, in contrast to cross-sectional data, panel data contain enough information to identify most, if not all, parameters in a typical measurement error model. Instead of using traditional econometric methods to analyze these models, the moment structure analysis approach was taken. That is, identification and estimation of model parameters was considered given only information about the first two moments of the data. For the random effect errors-in-variables model discussed in the first paper and the random coefficient model considered in the second paper of this dissertation, most parameters were shown to be identified given only information about these first two moments and relatively unrestrictive conditions. Aside from this, an estimation method that utilizes only the first two moments of the data was also proposed. Although such a method was originally developed for normally distributed data, simulation results showed that it performs well even with non-normal data and yield valid results. Moreover, the proposed method is easy to implement using existing statistical packages that perform moment structure analysis. Thus, moment structure analysis proves to be a good approach to fitting random effect and random coefficient models in which there is measurement error in the explanatory variables.

The use of moment structure analysis as an approach to the identification and estimation of measurement error models for panel data opens up many possible areas for further research. For one, hypothesis tests regarding the parameters of interest in the model will have to be refined. The methods examined in this dissertation also assume that one has balanced data or that missing values are missing at random. In the social sciences, data are more likely to be observational in nature, rather than experimental. Thus, attrition in panel data surveys may be a concern. Methods for measurement error panel data models for unbalanced data are therefore also needed.



## APPENDIX 1: ADDITIONAL SIMULATION RESULTS FOR THE RANDOM EFFECT ERRORS-IN-VARIABLES MODEL

A simulation study was also conducted using the random effects errors-in-variables model with two explanatory variables measured with error, given by

$$\mathbf{Y}_i = \alpha_i \mathbf{1}_T + \begin{pmatrix} \mathbf{x}_i^{(1)} & \mathbf{x}_i^{(2)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \mathbf{e}_i, \quad (1a)$$

$$\begin{pmatrix} \mathbf{X}_i^{(1)} \\ \mathbf{X}_i^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^{(1)} \\ \mathbf{x}_i^{(2)} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_i^{(1)} \\ \mathbf{u}_i^{(2)} \end{pmatrix}, \quad (1b)$$

$$\mathbf{E} \left[ \begin{pmatrix} \alpha_i \\ \mathbf{x}_i^{(1)} \\ \mathbf{x}_i^{(2)} \end{pmatrix} \right] = \begin{pmatrix} \mu_\alpha \\ \mu_{x_1} \\ \mu_{x_2} \end{pmatrix}, \quad (1c)$$

$$\mathbf{V} \left[ \begin{pmatrix} \alpha_i \\ \mathbf{x}_i^{(1)} \\ \mathbf{x}_i^{(2)} \end{pmatrix} \right] = \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma'_{x_1\alpha} & \sigma'_{x_2\alpha} \\ \sigma_{x_1\alpha} & \Sigma_{x_1x_1} & \Sigma'_{x_2x_1} \\ \sigma_{x_2\alpha} & \Sigma_{x_2x_1} & \Sigma_{x_2x_2} \end{bmatrix}, \quad (1d)$$

$$\mathbf{E}[\mathbf{e}_i] = \mathbf{0}, \quad \mathbf{E}[\mathbf{u}_i] = \mathbf{0}, \quad (1e)$$

$$\mathbf{V}[\mathbf{e}_i] = \Sigma_{ee}, \quad (1f)$$

$$\mathbf{V} \begin{bmatrix} \mathbf{u}_i^{(1)} \\ \mathbf{u}_i^{(2)} \end{bmatrix} = \begin{bmatrix} \Sigma_{uu_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{uu_2} \end{bmatrix}, \quad (1g)$$

where  $\mathbf{x}_i^{(j)} = (x_{i1}^{(j)}, \dots, x_{iT}^{(j)})'$  and  $\mathbf{u}_i^{(j)} = (u_{i1}^{(j)}, \dots, u_{iT}^{(j)})'$  for  $j = 1, 2$ ,  $i = 1, 2, \dots, n$ . The following sections contain results for simulations using model (1) with  $T = 3$  and  $n = 200$ . The first section compares the proposed pseudo maximum normal likelihood (ML) estimator with the Griliches and Hausman (1986) instrumental variable estimator (IV/GMM) of  $\beta = (\beta_1, \beta_2)'$  when observations are normally distributed. The second section assesses the finite sample properties of the inference procedures based on the pseudo maximum normal likelihood method when observations are not normally distributed. Throughout, ML estimates were computed using the SAS CALIS procedure, and the

Cholesky decomposition parameterization was used for any unrestricted covariance matrix.

#### IV/GMM Estimation versus ML Estimation

Data were generated according to model (1) in which all random variables, including  $\alpha_i$ ,  $x_{it}$ ,  $u_{it}$  and  $e_{it}$ , were normally distributed. The true values of the parameters were set to

$$\begin{aligned} \beta_1 = 1, \quad \beta_2 = 2, \quad \mu_\alpha = 4, \quad \sigma_{\alpha\alpha} = 1, \quad \sigma_{x_1\alpha} = \sigma_{x_2\alpha} = 0, \\ \Sigma_{uu_1} = \Sigma_{uu_2} = I_3, \quad \text{and} \quad \Sigma_{ee} = I_3. \end{aligned} \quad (2)$$

Two sets of true values for  $\mu_x = (\mu'_{x_1}, \mu'_{x_2})'$  and  $\Sigma_{xx}$ , described later, were considered. For the ML estimation, model (1), with  $\sigma_{x\alpha} = (\sigma'_{x_1\alpha}, \sigma'_{x_2\alpha})'$  estimated,  $\Sigma_{ee} = \sigma_{ee}I_3$ , and

$$\Sigma_{uu_j} = \begin{bmatrix} \sigma_{11}^{(j)} & \sigma_{21}^{(j)} & 0 \\ \sigma_{21}^{(j)} & \sigma_{22}^{(j)} & \sigma_{32}^{(j)} \\ 0 & \sigma_{32}^{(j)} & \sigma_{33}^{(j)} \end{bmatrix}, \quad j = 1, 2, \quad (3)$$

was fitted to each sample via maximum normal likelihood. The Griliches and Hausman (1986) IV/GMM estimator of  $\beta$  was also computed. Throughout, instruments listed in Table 1 were used. These instruments would all be valid under the assumed knowledge that for  $j = 1, 2$ ,

$$V[x_{i1}^{(j)}] + \{E[x_{i1}^{(j)}]\}^2 \neq V[x_{i2}^{(j)}] + \{E[x_{i2}^{(j)}]\}^2, \quad (4a)$$

$$\text{Cov}[x_{i1}^{(j)}, x_{i3}^{(j)}] + E[x_{i1}^{(j)}]E[x_{i3}^{(j)}] \neq \text{Cov}[x_{i2}^{(j)}, x_{i3}^{(j)}] + E[x_{i2}^{(j)}]E[x_{i3}^{(j)}], \quad (4b)$$

$$V[x_{i2}^{(j)}] + \{E[x_{i2}^{(j)}]\}^2 \neq V[x_{i3}^{(j)}] + \{E[x_{i3}^{(j)}]\}^2, \quad (4c)$$

$$\text{Cov}[x_{i1}^{(j)}, x_{i2}^{(j)}] + E[x_{i1}^{(j)}]E[x_{i2}^{(j)}] \neq \text{Cov}[x_{i1}^{(j)}, x_{i3}^{(j)}] + E[x_{i1}^{(j)}]E[x_{i3}^{(j)}], \quad (4d)$$

$$V[x_{i1}^{(j)}] + \{E[x_{i1}^{(j)}]\}^2 \neq V[x_{i3}^{(j)}] + \{E[x_{i3}^{(j)}]\}^2, \quad (4e)$$

$$\text{Cov}[x_{i1}^{(j)}, x_{i2}^{(j)}] + E[x_{i1}^{(j)}]E[x_{i2}^{(j)}] \neq \text{Cov}[x_{i2}^{(j)}, x_{i3}^{(j)}] + E[x_{i2}^{(j)}]E[x_{i3}^{(j)}]. \quad (4f)$$

In addition to the  $\tilde{\beta}_{\text{GMM}}$  defined by Griliches and Hausman (1986), a modified version of  $\tilde{\beta}$ , denoted by  $\tilde{\beta}_{\text{GMM}}^*$ , was computed. The modification involved using a more efficient weight matrix arising from the assumption that both  $e_{it}$  and  $u_{it}^{(j)}$  are independent and identically distributed over  $i$  and  $t$ .

The first set of 1000 samples was generated according to model (1) with true parameter values (2) and, in addition,

$$\mu_{x_1} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mu_{x_2} = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \Sigma_{xx} = \begin{bmatrix} \Sigma_{x_1x_1} & 0 \\ 0 & \Sigma_{x_2x_2} \end{bmatrix}, \quad (5a)$$

Table 1 Instruments used to compute initial estimates of  $\beta$ 

Equation to be estimated	Instruments
$Y_{i2} - Y_{i1} = \beta_1(x_{i2}^{(1)} - x_{i1}^{(1)}) + \beta_2(x_{i2}^{(2)} - x_{i1}^{(2)}) + (\epsilon_{i2} - \epsilon_{i1})$	$X_{i1}^{(1)} + X_{i2}^{(1)}, X_{i3}^{(1)};$ $X_{i1}^{(2)} + X_{i2}^{(2)}, X_{i3}^{(2)}$
$Y_{i3} - Y_{i2} = \beta_1(x_{i3}^{(1)} - x_{i2}^{(1)}) + \beta_2(x_{i3}^{(2)} - x_{i2}^{(2)}) + (\epsilon_{i3} - \epsilon_{i2})$	$2(X_{i2}^{(1)} + X_{i3}^{(1)}), X_{i1}^{(1)};$ $2(X_{i2}^{(2)} + X_{i3}^{(2)}), X_{i1}^{(2)}$
$Y_{i3} - Y_{i1} = \beta_1(x_{i3}^{(1)} - x_{i1}^{(1)}) + \beta_2(x_{i3}^{(2)} - x_{i1}^{(2)}) + (\epsilon_{i3} - \epsilon_{i1})$	$X_{i1}^{(1)} + X_{i3}^{(1)}, X_{i2}^{(1)};$ $X_{i1}^{(2)} + X_{i3}^{(2)}, X_{i2}^{(2)}$

where

$$\Sigma_{x_1x_1} = \begin{bmatrix} 4.0 & 0.2 & 3.2 \\ 0.2 & 1.0 & 0.2 \\ 3.2 & 0.2 & 5.0 \end{bmatrix}, \quad \text{and} \quad \Sigma_{x_2x_2} = \begin{bmatrix} 2.0 & 0.2 & 0.1 \\ 0.2 & 10.0 & 0.1 \\ 0.1 & 0.1 & 6.0 \end{bmatrix}. \quad (5b)$$

The ML and IV/GMM estimates of  $\beta$  were computed in the manner described above. Figure 1 presents box plots of the resulting 1000 estimates. The bias, variance and mean squared error for the three estimators are given in Table 2. For the true values (5), conditions (4a) and (4b) are not satisfied. In other words, the instruments used for the first equation are invalid. As a result, the IV/GMM and modified IV/GMM estimators of  $\beta$  have large bias. Hence, the performance of the IV/GMM estimator can be unsatisfactory when specific knowledge assumed for the instrument choice is incorrect. On the other hand, the ML estimator has smaller bias despite the fact that more parameters are estimated together with  $\beta$  (e.g.,  $\sigma_{x\alpha}$  and  $\Sigma_{uu}, j = 1, 2$ , of the form (3)).

Table 2 Monte Carlo moments for case 1

Estimator	Bias	Variance	MSE
$\tilde{\beta}_{1,\text{GMM}}$	-0.03195	0.01396	0.01498
$\tilde{\beta}_{2,\text{GMM}}$	-0.03083	0.00753	0.00848
$\tilde{\beta}_{1,\text{GMM}}^*$	-0.03242	0.01334	0.01439
$\tilde{\beta}_{2,\text{GMM}}^*$	-0.02975	0.00715	0.00803
$\hat{\beta}_{1,\text{ML}}$	-0.02128	0.01939	0.01984
$\hat{\beta}_{2,\text{ML}}$	-0.01116	0.00618	0.00630

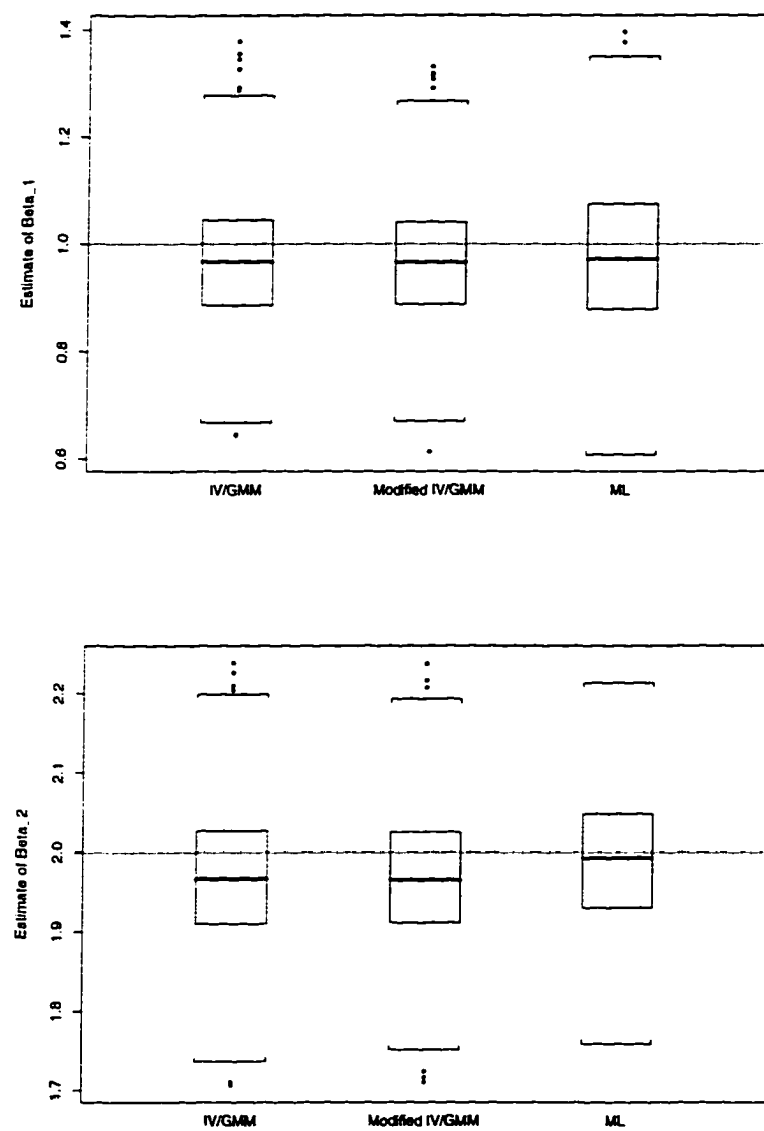


Figure 1 Box plots of estimates of  $\beta_1$  and  $\beta_2$  for case 1

The second set of 1000 samples was generated using the true values (2) and

$$\mu_{x_1} = \begin{pmatrix} 2 \\ 5 \\ 10 \end{pmatrix}, \quad \mu_{x_2} = \begin{pmatrix} 3 \\ 5 \\ 15 \end{pmatrix}, \quad \text{and} \quad \Sigma_{xx} = \begin{bmatrix} \Sigma_{x_1x_1} & 0 \\ 0 & \Sigma_{x_2x_2} \end{bmatrix}. \quad (6a)$$

where

$$\Sigma_{x_1x_1} = \begin{bmatrix} 4.0 & 0.5 & 1.0 \\ 0.5 & 4.8 & 0.8 \\ 1.0 & 0.8 & 5.6 \end{bmatrix}, \quad \text{and} \quad \Sigma_{x_2x_2} = \begin{bmatrix} 5.0 & 0.5 & 0.2 \\ 0.5 & 5.5 & 0.7 \\ 0.2 & 0.7 & 6.5 \end{bmatrix}. \quad (6b)$$

In this case, all the knowledge assumed for the IV/GMM estimation is in fact correct, and all instruments in Table 1 are valid. As shown in the simulation results given in Table 3 and Figure 2, the IV/GMM and the modified IV/GMM estimators using informative instruments perform as efficiently as the ML estimator. Even in such a situation, the ML method that estimates more parameters under more general conditions produces an estimator of  $\beta$  with comparable bias and variance as the IV/GMM estimators.

Table 3 Monte Carlo moments for case 2

Estimator	Bias	Variance	MSE
$\tilde{\beta}_{1,\text{GMM}}$	0.00628	0.01600	0.01603
$\tilde{\beta}_{2,\text{GMM}}$	-0.00417	0.00631	0.00633
$\tilde{\beta}_{1,\text{GMM}}^*$	0.00671	0.01538	0.01542
$\tilde{\beta}_{2,\text{GMM}}^*$	-0.00456	0.00608	0.00610
$\hat{\beta}_{1,\text{ML}}$	0.01751	0.01647	0.01678
$\hat{\beta}_{2,\text{ML}}$	-0.00999	0.00645	0.00655

## Finite Sample Properties of the ML Estimator

Simulations were also conducted to assess the finite sample properties of the maximum normal likelihood estimator when observations are not normal. The following set of results deals with situations in which the equation and measurement errors are normal but the true  $x$ 's and the random effect  $\alpha_i$  are non-normal. By Theorem 4, the standard error estimates and chi-square tests of model fit obtained under the normality of observations are still valid, although observations are not normal. Data were generated according to model (1) with  $T = 3$ ,  $n = 200$ , and true parameter values (2) and (6). The equation errors  $e_{it}$  and the measurement errors  $u_{it}^{(j)}$ ,  $j = 1, 2$ , were normally distributed. Four different

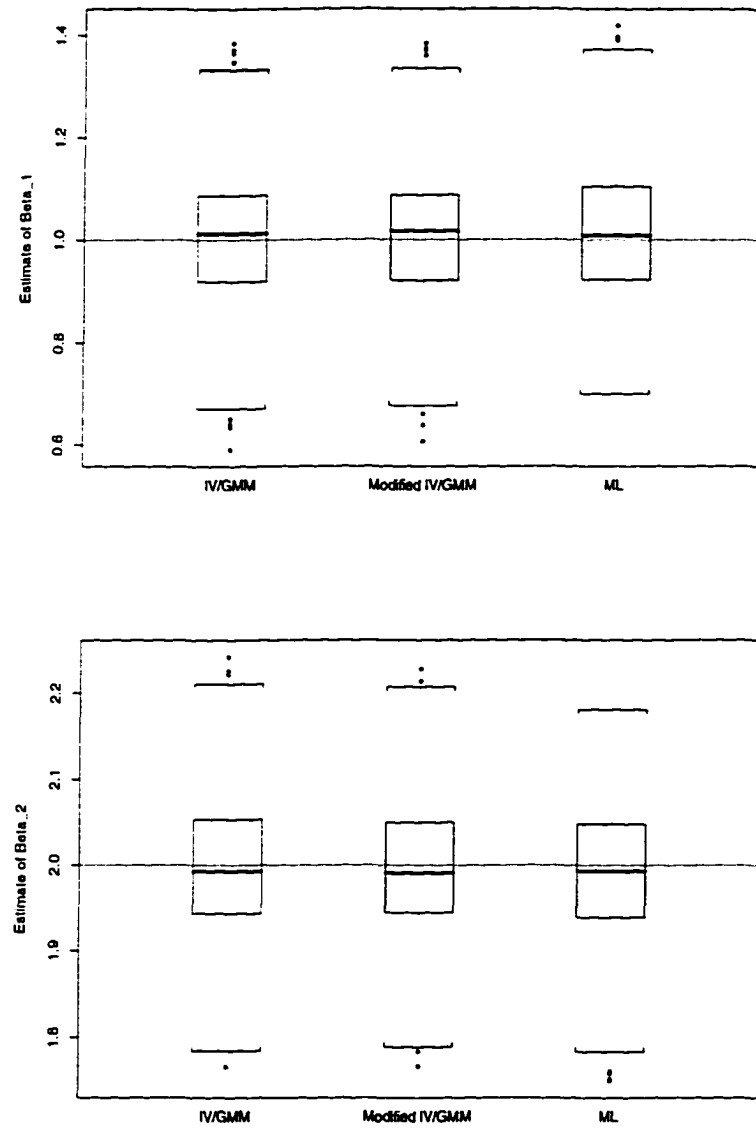


Figure 2 Box plots of estimates of  $\beta_1$  and  $\beta_2$  for case 2

sets of 1000 samples were generated with different distributions of  $x_{it}$  and  $\alpha_i$  listed in Tables 4 and 6, where N indicates a normal distribution and  $\chi^2$  indicates a shifted and scaled  $\chi^2$  distribution with two degrees of freedom. Note that the first set with normal  $x_i$  and normal  $\alpha_i$  actually gives normal observations. Two versions of model (1) were fitted to each sample via maximum normal likelihood. One version, with  $\sigma_{x\alpha} = 0$  not estimated, involved fitting an unrestricted  $3 \times 3$  matrix for each  $\Sigma_{uu_j}$ ,  $j = 1, 2$ , and fitting  $\Sigma_{ee}$  as  $\sigma_{ee}\mathbf{I}_3$ . In the second version,  $\sigma_{x\alpha}$  was estimated,  $\Sigma_{ee}$  was fitted as  $\sigma_{ee}\mathbf{I}_3$ , and each  $\Sigma_{uu_j}$  was fitted as  $\sigma_{uu_j}\mathbf{I}_3$ ,  $j = 1, 2$ . The goodness-of-fit statistic  $F(\hat{\theta})$  as well as 95% confidence intervals for  $\mu_\alpha$ ,  $\beta_1$  and  $\beta_2$  under pseudo normality were obtained for each sample and for each method of fitting model (1). Simulated coverage probabilities are reported in Table 4 while simulated probabilities for  $F(\hat{\theta})$  to exceed the 0.10, 0.05 and 0.01 upper quantiles of the appropriate  $\chi^2$  distribution are shown in Table 5. The coverage probabilities in Table 4 are reasonably similar to the nominal value of 0.95. Thus, valid inferences about  $\mu_\alpha$ ,  $\beta_1$  and  $\beta_2$  can be made using standard errors computed under the pseudo normality of observations for non-normal distributions of the random effect  $\alpha_i$  and the true explanatory variables  $x_i$  when the equation and measurement errors are normally distributed. However, as shown in Table 5, the upper percentiles of the fit statistic are not well approximated by the  $\chi^2$  percentiles. Many of the large goodness-of-fit statistics were found to be associated with samples for which a near-singular estimate of  $\Sigma_{xx}$  was obtained. An attempt to estimate model (1) while at the same time avoiding getting near-singular estimates of  $\Sigma_{xx}$  was made. This involved adding a small fraction (say,  $\frac{1}{n}$ ) of the sample covariance matrix of  $\text{vec } x_i$  to the estimate of  $\Sigma_{xx}$  then re-estimating model (1) via ML given this adjusted  $\Sigma_{xx}$  estimate. The upper percentiles of the resulting goodness-of-fit statistics were still not well approximated by the  $\chi^2$  percentiles.

Table 4 Simulated coverage probabilities of 95% confidence intervals  
(normal errors)

Error variances	Distribution		$\sigma_{x\alpha} = 0$			$\sigma_{x\alpha}$ estimated		
	$x_i$	$\alpha_i$	$\mu_\alpha$	$\beta_1$	$\beta_2$	$\mu_\alpha$	$\beta_1$	$\beta_2$
$\sigma_{ee} = \sigma_{uu} = 1$ :								
	N	N	0.965	0.972	0.977	0.946	0.900	0.892
	$\chi^2$	N	0.967	0.973	0.970	0.952	0.919	0.916
	N	$\chi^2$	0.970	0.966	0.967	0.954	0.906	0.913
	$\chi^2$	$\chi^2$	0.973	0.972	0.976	0.951	0.913	0.908
$\sigma_{ee} = \sigma_{uu} = 0.1$ :								
	N	N	0.963	0.996	0.998	0.937	0.937	0.961
	$\chi^2$	N	0.960	0.996	0.995	0.942	0.948	0.953
	N	$\chi^2$	0.962	0.995	0.996	0.940	0.943	0.962
	$\chi^2$	$\chi^2$	0.959	0.996	0.996	0.932	0.935	0.950

Table 5 Simulated frequency of rejection of the  $\chi^2$  goodness-of-fit test  
(normal errors)

Error variances	Distribution		$\sigma_{x\alpha} = 0$			$\sigma_{x\alpha}$ estimated		
	$x_i$	$\alpha_i$	10%	5%	1%	10%	5%	1%
$\sigma_{ee} = \sigma_{uu} = 1$ :								
	N	N	0.373	0.334	0.292	0.322	0.233	0.138
	$\chi^2$	N	0.397	0.347	0.308	0.330	0.244	0.124
	N	$\chi^2$	0.349	0.301	0.276	0.288	0.216	0.119
	$\chi^2$	$\chi^2$	0.390	0.345	0.314	0.341	0.261	0.146
$\sigma_{ee} = \sigma_{uu} = 0.1$ :								
	N	N	0.275	0.193	0.117	0.270	0.230	0.179
	$\chi^2$	N	0.250	0.180	0.100	0.260	0.213	0.163
	N	$\chi^2$	0.273	0.191	0.117	0.269	0.234	0.199
	$\chi^2$	$\chi^2$	0.148	0.112	0.090	0.251	0.201	0.166



## APPENDIX 2: ADDITIONAL TABLES FOR RANDOM COEFFICIENT REGRESSION WITH ERRORS IN VARIABLES

The following tables present additional results of the simulation study on the finite-sample behavior of GLS and PML estimators for the random coefficient regression model with errors in variables, given by

$$\mathbf{Y}_i = \begin{pmatrix} 1 & \mathbf{X}_i \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + e_i, \quad (1a)$$

$$\mathbf{X}_i = \mathbf{x}_i + \mathbf{u}_i, \quad (1b)$$

$$E \left[ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \right] = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad (1c)$$

$$\Phi = \begin{bmatrix} \phi_{\alpha\alpha} & \phi_{\beta\alpha} \\ \phi_{\beta\alpha} & \phi_{\beta\beta} \end{bmatrix}. \quad (1d)$$

The estimates reported here are the result of fitting model (1) using GLS and using PML to each of 1000 samples of size  $n = 50, 200$ , and  $500$  with normally distributed  $(\alpha_i, \beta_i)$ ,  $\mathbf{x}_i$ ,  $\mathbf{u}_i$  and  $e_i$  and with  $\text{var}[e_i] = \sigma_{ee}\mathbf{I}_3$  and  $\text{var}[\mathbf{u}_i] = \sigma_{uu}\mathbf{I}_3$ . Tables 1 to 3 report the bias and mean squared error of the GLS and PML estimators. Tables 4 and 5 present the simulation variances of the estimators and the average of the estimated variances for the cases with  $n = 50$  and  $n = 200$ .<sup>1</sup>

---

<sup>1</sup> A table presenting variance estimates for the case with  $n = 500$  is given on page 53.

Table 1 Bias and mean squared errors of estimators ( $n = 50$ )

Parameter	True Value	Bias		MSE	
		GLS	PML	GLS	PML
$\mu_\alpha$	4.0	-0.8615	-0.0160	2.4116	0.5601
$\mu_\beta$	1.0	-0.0202	0.0153	0.0634	0.0831
$\phi_{\alpha\alpha}$	2.0	2.4607	2.7797	40.3986	43.3616
$\phi_{\beta\alpha}$	1.0	-0.6808	-0.4211	3.1874	3.8314
$\phi_{\beta\beta}$	3.0	-0.2648	-0.0799	1.0729	0.6173
$\mu_{x_1}$	2.0	-0.0268	-0.0072	0.1888	0.1507
$\mu_{x_2}$	5.0	-0.0646	-0.0142	0.2148	0.1679
$\mu_{x_3}$	10.0	-0.0465	0.0100	0.2213	0.1856
$\sigma_{x_1, x_1}$	4.0	0.9596	-0.2867	7.7029	2.5655
$\sigma_{x_2, x_1}$	2.0	-0.9739	-0.1932	2.0306	1.1120
$\sigma_{x_3, x_1}$	0.8	-0.1904	0.0510	1.1579	1.2545
$\sigma_{x_2, x_2}$	4.8	0.9743	-0.1980	6.8817	3.1097
$\sigma_{x_3, x_2}$	2.8	-1.0436	-0.0645	3.2779	1.4188
$\sigma_{x_3, x_3}$	5.6	1.1774	-0.2981	8.0164	4.3076
$\sigma_{uu}$	4.0	-1.3087	0.0337	4.4989	1.6216
$\sigma_{ee}$	4.0	-1.5015	-0.3841	11.5000	10.8481

Table 2 Bias and mean squared errors of estimators ( $n = 200$ )

Parameter	True Value	Bias		MSE	
		GLS	PML	GLS	PML
$\mu_\alpha$	4.0	-0.2243	-0.0323	0.5035	0.1487
$\mu_\beta$	1.0	0.0033	0.0151	0.0167	0.0196
$\phi_{\alpha\alpha}$	2.0	0.8938	0.6635	11.4675	11.3131
$\phi_{\beta\alpha}$	1.0	-0.3740	-0.0722	1.0940	0.9680
$\phi_{\beta\beta}$	3.0	0.2881	-0.0149	0.9701	0.1597
$\mu_{x_1}$	2.0	-0.0027	-0.0030	0.0412	0.0381
$\mu_{x_2}$	5.0	-0.0172	-0.0036	0.0397	0.0397
$\mu_{x_3}$	10.0	-0.0068	0.0133	0.0483	0.0457
$\sigma_{x_1, x_1}$	4.0	0.2801	-0.1590	1.7743	0.7490
$\sigma_{x_2, x_1}$	2.0	-0.3244	-0.0366	0.4901	0.2947
$\sigma_{x_3, x_1}$	0.8	-0.0588	0.0481	0.3678	0.3612
$\sigma_{x_2, x_2}$	4.8	0.3720	-0.0856	1.7343	0.8354
$\sigma_{x_3, x_2}$	2.8	-0.3073	0.0407	0.8106	0.4182
$\sigma_{x_3, x_3}$	5.6	0.5018	-0.0810	2.4701	1.2150
$\sigma_{uu}$	4.0	-0.0691	0.0639	1.8095	0.4455
$\sigma_{ee}$	4.0	-0.8457	0.1575	6.6618	4.9999

Table 3 Bias and mean squared errors of estimators ( $n = 500$ )

Parameter	True Value	Bias		MSE	
		GLS	PML	GLS	PML
$\mu_\alpha$	4.0	-0.0314	0.0044	0.0967	0.0626
$\mu_\beta$	1.0	-0.0018	0.0042	0.0070	0.0077
$\phi_{\alpha\alpha}$	2.0	0.4993	0.0639	6.3431	6.1122
$\phi_{\beta\alpha}$	1.0	-0.2976	-0.0190	0.5526	0.4061
$\phi_{\beta\beta}$	3.0	0.3193	-0.0164	0.7284	0.0738
$\mu_{x_1}$	2.0	-0.0042	-0.0027	0.0161	0.0159
$\mu_{x_2}$	5.0	-0.0034	0.0003	0.0162	0.0163
$\mu_{x_3}$	10.0	-0.0019	0.0079	0.0189	0.0184
$\sigma_{x_1, x_1}$	4.0	0.0637	-0.0496	0.4627	0.3052
$\sigma_{x_2, x_1}$	2.0	-0.1292	-0.0025	0.1565	0.1168
$\sigma_{x_3, x_1}$	0.8	-0.0459	-0.0143	0.1443	0.1408
$\sigma_{x_2, x_2}$	4.8	0.1379	-0.0409	0.5391	0.3681
$\sigma_{x_3, x_2}$	2.8	-0.1373	-0.0132	0.2555	0.1731
$\sigma_{x_3, x_3}$	5.6	0.1690	-0.0552	0.7842	0.4890
$\sigma_{uu}$	4.0	0.0685	0.0029	0.6758	0.1627
$\sigma_{ee}$	4.0	-0.1250	0.1213	5.6300	2.2579

Table 4 Variance estimates for parameters ( $n = 50$ )

Parameter	GLS		PML	
	Simulation variance	Average estimated variance <sup>a</sup>	Simulation variance	Average estimated variance
$\alpha$	1.6694	0.2860	0.5599	0.5812
$\beta$	0.0630	0.0428	0.0829	0.0753
$\phi_{\alpha\alpha}$	34.3436	$1.08 \times 10^8$	4.2529	91.8827
$\phi_{\beta\alpha}$	2.7239	$1.26 \times 10^{14}$	0.6418	4.4165
$\phi_{\beta\beta}$	1.0028	$4.33 \times 10^8$	0.5791	0.6641
$\mu_{x_1}$	0.1881	0.1062	0.1507	0.1519
$\mu_{x_2}$	0.2107	0.1003	0.1677	0.1519
$\mu_{x_3}$	0.2191	0.1356	0.1855	0.1847
$\sigma_{x_1, x_1}$	6.7820	48.9146	3.7153	2.9684
$\sigma_{x_2, x_1}$	1.0822	$1.94 \times 10^6$	1.1704	1.1285
$\sigma_{x_3, x_1}$	1.1217	$1.59 \times 10^{11}$	2.2880	1.2507
$\sigma_{x_2, x_2}$	5.9324	$1.52 \times 10^5$	0.5852	3.3610
$\sigma_{x_3, x_2}$	2.1889	$3.53 \times 10^{12}$	1.1897	1.5641
$\sigma_{x_3, x_3}$	6.6301	$6.12 \times 10^{10}$	2.7991	4.8336
$\sigma_{uu}$	2.7862	1.9983	1.6205	1.8441
$\sigma_{ee}$	9.2457	22.8906	10.7006	26.7966

<sup>a</sup>Median GLS variance estimates for  $\phi_{\alpha\alpha}$ ,  $\phi_{\beta\alpha}$  and  $\phi_{\beta\beta}$  were 56.3758, 3879943, and 131.8147, respectively.

Table 5 Variance estimates for parameters ( $n = 200$ )

Parameter	GLS		PML	
	Simulation variance	Average estimated variance <sup>a</sup>	Simulation variance	Average estimated variance
$\alpha$	0.4532	0.0949	0.1477	0.1462
$\beta$	0.0167	0.0154	0.0194	0.0197
$\phi_{\alpha\alpha}$	10.6686	$1.27 \times 10^{10}$	10.8728	24.0378
$\phi_{\beta\alpha}$	0.9541	$3.06 \times 10^{13}$	0.9627	1.1175
$\phi_{\beta\beta}$	0.8871	$5.09 \times 10^{10}$	0.1595	0.1682
$\mu_{x_1}$	0.0412	0.0343	0.0380	0.0378
$\mu_{x_2}$	0.0395	0.0333	0.0397	0.0397
$\mu_{x_3}$	0.0482	0.0438	0.0455	0.0471
$\sigma_{x_1, x_1}$	1.6958	0.7217	0.7237	0.6685
$\sigma_{x_2, x_1}$	0.3848	0.2971	0.2934	0.2987
$\sigma_{x_3, x_1}$	0.3644	0.3406	0.3589	0.3322
$\sigma_{x_2, x_2}$	1.5959	0.8303	0.8281	0.7852
$\sigma_{x_3, x_2}$	0.7162	0.5998	0.4166	0.4167
$\sigma_{x_3, x_3}$	2.2183	0.9479	1.2085	1.1689
$\sigma_{uu}$	1.8048	0.3502	0.4414	0.3785
$\sigma_{ee}$	5.9466	6.4391	4.9751	5.7484

<sup>a</sup>Median GLS variance estimates for  $\phi_{\alpha\alpha}$ ,  $\phi_{\beta\alpha}$  and  $\phi_{\beta\beta}$  were 22.2383, 1.5589, and 5.7510, respectively.

## ACKNOWLEDGMENTS

I would like to express my sincere thanks to my major professors, Dr. Yasuo Amemiya and Dr. Arne Hallam, for their guidance throughout my graduate education. I benefited greatly from discussions with Dr. Amemiya and from his expertise on measurement error models and structural equation modeling. I also acknowledge assistance from other members of my program of study committee: Dr. Peter Orazem, Dr. Jay Breidt, Dr. Tahira Hira and Dr. Doug Miller.

To all the professors at Iowa State who contributed to my education in statistics and economics, I wish to express my gratitude. I thank Dr. Arne Hallam and Dr. Peter Orazem for generously funding my stay with the Department of Economics. I am deeply indebted, in particular, to Dr. Frances Antonovitz, Dr. Arne Hallam, Dr. Peter Orazem and Dr. Leigh Tesfatsion for encouraging me to continue pursuing a doctorate in economics even when I started to doubt my ability to do so.

I am grateful for the love and support of all my friends and family. Last but not the least, I thank the Lord for His guidance in all my endeavors and for giving meaning in everything I do.